

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:35:00 ; Search time 2025.8 Seconds
(without alignments)
16520.996 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1352_2501

Perfect score: 1150
Sequence: 1 cgggtaacaagcagcgagac.....attgaggagagcgaagcga 1150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sy:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_higo_hum:*
- 40: em_higo_mus:*
- 41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1150	100.0	3442	9	AB008430	AB008430 Homo sapi
2	879.4	76.5	3902	10	BC030329	BC030329 Mus muscu
3	478.8	41.6	2632	10	BC004009	BC004009 Mus muscu
4	335.6	29.2	3719	10	BC009153	BC009153 Mus muscu
5	329.6	28.7	3997	9	AB018336	AB018336 Homo sapi
6	312.8	27.2	2431	10	BC027077	BC027077 Mus muscu
7	241	21.0	106578	9	AL137249	AL137249 Human DNA
8	204.2	17.8	224310	2	AC094777	AC094777 Rattus no
9	199.4	17.3	221502	2	AC122885	AC122885 Mus muscu
10	193.4	16.8	152053	2	AC022669	AC022669 Homo sapi
11	162.4	14.1	393	11	G60059	G60059 SHGC-130885
12	154.6	13.4	426	6	AX261839	AX261839 Sequence
13	150.2	13.1	120911	2	AC130919	AC130919 Rattus no
14	148.6	12.9	190588	2	AC101725	AC101725 Mus muscu
15	148.6	12.9	221502	2	AC122885	AC122885 Mus muscu
16	122.6	10.7	2605	9	BC021301	BC021301 Homo sapi
17	112	9.7	96183	9	AL161896	AL161896 Human DNA
18	112	9.7	152053	2	AC022669	AC022669 Homo sapi
19	111.8	9.7	149417	2	AC118779	AC118779 Rattus no
20	105.8	9.2	42014	2	AC018327	AC018327 Drosophill
21	105.8	9.2	160710	3	AC009537	AC009537 Drosophill
22	105.8	9.2	164713	3	AE007589	AE007589 Drosophill
23	105.8	9.2	303209	3	AE003604	AE003604 Drosophill
24	97	8.4	62537	2	AC101100	AC101100 Mus muscu
25	95.8	8.3	105605	2	AC115336	AC115336 Rattus no
26	95.8	8.3	176630	2	AC109427	AC109427 Rattus no
27	95	8.3	3055	10	BC026860	BC026860 Mus muscu
28	87.8	7.6	134911	9	AC005104	AC005104 Homo sapi
29	83.6	7.3	1296	6	AX411309	AX411309 Sequence
30	83.6	7.3	1821	9	AK002045	AK002045 Homo sapi
31	83.6	7.3	3610	6	AX405929	AX405929 Sequence
32	83.6	7.3	3842	9	AB037783	AB037783 Homo sapi
33	82	7.1	1478	9	BC013319	BC013319 Homo sapi
34	81.4	7.1	3220	5	AF017370	AF017370 Danio rer
35	80.2	7.0	67896	2	AC131316	AC131316 Mus muscu
36	79.6	6.9	133745	2	AC131454	AC131454 Strongylo
37	72.8	6.3	462	6	AX371287	AX371287 Sequence
38	70.8	6.2	138025	9	HSBJ20N4	AL096709 Human DNA
39	68	5.9	149417	2	AC118779	AC118779 Rattus no
40	65.8	5.7	2956	10	AF017369	AF017369 Mus muscu
41	64.4	5.6	1896	9	BC032232	BC032232 Homo sapi
42	64.4	5.6	3224	9	BC032429	BC032429 Homo sapi
43	64.4	5.6	4670	9	AK000004	AK000004 Homo sapi
44	62.2	5.4	2170	9	AK097217	AK097217 Homo sapi
45	62.2	5.4	120911	2	AC130919	AC130919 Rattus no

ALIGNMENTS

RESULT 1
AB008430
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AB008430
Homo sapiens mRNA for CDEP, complete cds.
AB008430.1
GI:2766164
CDEP.
Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and
Kato,Y.
Molecular cloning and characterization of CDEP, a novel human

3442 bp
mRNA
linear
PRI 13-FEB-1999

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors
Biochem Biophys. Res. Commun. 241 (2), 369-375 (1997)
98086358

2 (bases 1 to 3442)
Koyano,Y., Kawamoto,T. and Kato,Y.
Direct Submission
Submitted (22-Oct-1997) Takeshi Kawamoto, Hiroshima University
School of Dentistry, Department of Biochemistry, 1-2-3 Kasumi
Minami-ku, Hiroshima, Hiroshima 734, Japan
(E-mail:tkawano@ipc.hiroshima-u.ac.jp, Tel:082-257-5688,
Fax:082-257-5629)

FEATURES

source

Location/Qualifiers
1..3442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="chondrocyte"
/tissue_type="cartilage"
/dev_stage="embryo"
49..3186
/function="Rho Guanine Nucleotide Exchange Factor"
/note="Band 4.1 superfamily"

CDS

/codon_start=1
/product="CDEP"
/protein_id="BAA24267.1"
/db_xref="GI:2766165"
/translation="MGEIQRTPGSRILGAPENSIGSLERQKPPPTPSKLVSIKI
QMLDDQFAVEVQFQAPKQVLLDVAENHLNLVEGDFGLFDPDHRKRTITWDLKPIV
KOIRRRVHVYKVFVFFPDHDTQLEELTFALQVKQDLAQGRLLTCNDTSALLI
SHVQSEIGDDEADLEHLAKNKYIPQDDALELIVFEFHNNHIGQTPAESDFQLLEI
ARRLMYGTIRLHPAKDRGTINKLAVANTGILVFOGFTKINAFNWKVKLSFRKRF
LKLPRDANSAYQDFTLELMASRDFCKSFWKICVHHAFERFLFEPKPKPVLFSRG
SSEPSGTQKQVLYVEGKKVQFPERKSKIHISRLASQPIELSEVLEBQSQS
TSLTGEGAESGGSGRKEPVSAGEPSHPSPAPRSPAGNKADGASAPTEE
EEVYVDTQSQKPPQPPQSTGSLTSGPSHLSVNSQGGVAPANVTLSPLNPDTKQ
ASPLSLINDQAPCTDEDEGRKRFTDKAYFAKEVSTERTYTKLDELVTISWF
QSTVSKEDAMPALKSLIFPNPEPLHKFTNFKIEQRLALWEGRSNAQIRDYORIG
DVMKNIQGMKLAHLKHSFALENGIKSSRRLENFCRDFLOKVCVYLPNTFL
LPLRLHMYKQVLERCKHUHPSSHADPRDCAALAEJTEWVAQLHGMIMKWNFKL
HELKDGLIDNLVYVPGREFIKGLSKLSKGLQQRMFLLFNDVLLTSLRSTASNQ
FKVHGOLPIPIESEDWVPHCLTLRGQRQSIIVAASSRSMEKWKVEDIQMID
LAEKSSPAPEFLASDPNDKSPDEATAQDESDLSASRTSLEROAPHRGNTMHHV
CWHRTSVSWDFSTAVENQSLNLRKFKNSGWQKLVVTFNFCFLFYKSHQDNHP
LASPLGLYSLTIPSESENIQKDYVFKLHFKSHVYFRAESEYTERWMEVIRSATSS
ASRPHVLSKESLVI"

polya_site

3442

/note="50 a nucleotides"

BASE COUNT 864 a 952 c 927 g 699 t

ORIGIN

Query Match 100.0%; Score 1150; DB 9; Length 3442;

Best Local Similarity 100.0%; Pred. No. 6e-266;

Matches 1150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGTAACAGCAGCGGAGCGCGCTCGCGGCCCGCCAGGAGGAGGAGGTCG 60

Db 1352 CGGGTAACAGCAGCGGAGCGCGCTCGCGGCCCGCCAGGAGGAGGAGGTCG 1411

QY 61 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCCCGCAGCAGCAGGCTCCCTGA 120

Db 1412 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCCCGCAGCAGCAGGCTCCCTGA 1471

QY 121 CTGGAGTCTCACCCTTCGGAGCTCTCTGTAACCTCGCAGGGGGAGTGGCCCTGCCA 180

Db 1472 CTGGAGTCTCACCCTTCGGAGCTCTCTGTAACCTCGCAGGGGGAGTGGCCCTGCCA 1531

QY 181 ACGTGACCTTGTCTCCCACTCAGCGCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCC 240

Db 1532 ACGTGACCTTGTCTCCCACTCAGCGCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCC 1591

QY 241 CGTGTGTAATGACAGGCGTCCCGGACCGACGAGTGAAGGATAGGCGCGGAGGA 300

RESULT 2

BC030329

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC030329 3902 bp mRNA linear ROD 07-AUG-2002
Mus musculus, clone IMAGE:5376197, mRNA, partial cds.

BC030329
BC030329.1 GI:20987935

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db 1747 GACTGACGCATCTATCATGTTAAAGTCCAGGACAGCTCCCACTCTATGCGATGACGA 1806
QY 1132 TTGAGGAGAGCGAAGACGA 1150
Db 1807 TCAGGAGAGTGGAGGA 1825

RESULT 3
BC004009 2632 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone IMAGE:3493093, mRNA, partial cds.
DEFINITION BC004009
ACCESSION BC004009.1 GI:13278387
VERSION
KEYWORDS house mouse.
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2632)
Direct Submission
Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 8 Row: P Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
Location/Qualifiers
1..2632
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZECH II"
/clone="IMAGE:3493093"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMV."
/clonal_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..1263
/codon_start=1
/product="Unknown (protein for IMAGE:3493093)"
/protein_id="AAH04009.1"
/db_xref="GI:13278388"
/translation="VMLKNIQMKHLAAHLWKHSEALEALETSIKSRRLHFCRDFE
LQKCYLPNTFLRLPLRLHMYKHLERLCKHPHPNHADFRDCFAALAEITEMVAQL
HGTHLKMENFQKLEHLLKDLIGDNLVPGREFIRLGLSKLSKGLQORMFNFVNDV
LLYTSRGTAQNQKRGOLPLGYMTIEESEEWGVPCHLTLRGQRSDLIIVAASRSE
MEKMDTOMADIAEKNSGPTPELLASPPDKSPDPADESDDLASRTSLE
RQAPHRGNTVHVCHRSYVSMVDFSTAVENQLSNLLRFRKNSGNWQKWLWVFTNF
CLFFYKHQSDHPLASLPGLGYSLTIPSENIHKDYVFKLHFKSHVYFRAESEYTF
ERWMEVIRSATSSASRAHLSHKEHLY"
632 a 720 c 671 g 609 t

BASE COUNT
ORIGIN

Query Match 41.6%; Score 478.8; DB 10; Length 2632;

Best Local Similarity 89.3%; Pred. No. 1.8e-104;
Matches 516; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 573 GTCATGCTGAAGAACAATTCAGGGCATGAAGACACCTGGCGGCTCACTGTGGGAAGCACAGC 632
Db 1 GTCATGCTGAAGAACAATTCAGGGCATGAAGACACCTGGCGGCTCACTGTGGGAAGCACAGT 60
QY 633 GAGGCTTTGGAGGCGCTTGAGAAATCAAGAGCTCCCGGCGGTGGAGAACTTCTGC 692
Db 61 GAGGCGCTTGAGGCGCTTGAGAAATCAAGAGCTCCCGGCGGTGGAGAACTTCTGC 120
QY 693 AGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGCGCA 752
Db 121 CGAGACTTCGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGCGG 180
QY 753 CTGACCGGCTCATGCACTACAGCAGGTCTCTGGAGCGCTGTGCAAAACACACCGCGCG 812
Db 181 CTGACCGGCTCATGCACTATTAAGCATGTCTCTGGAGAGCTGTGCAAGCACACCAACCA 240
QY 813 AGCCAGCGGCTTCAGGAGCTGCCGAGCTGCGGAGCGCTTTGGCAGAGATCAGGAGATGTCGCA 872
Db 241 AACCCAGCGGCTTCAGGAGCTGCGGAGCTGCGGCGGAGATCAGAGATGTCGCGC 300
QY 873 CAGCTCCAGCGGTACGATGATCAAGATGGAAGATTTCCAGAAAGCTGCACGAAGCTCAAGAAA 932
Db 301 CAGCTGCAGCGGACCATGATCAAGATGGAAGATTTCCAGAAAGCTGCATGAGCTCAAGAA 360
QY 933 GATTGATTGGCATTTGACAAATCTTGTGGTCCGGAAGGAGTTTCATCGTCTGGGCGAG 992
Db 361 GATGCTCGGCAATGACAAATCTTGTGATCCAGGAAGGAGTTTCATCGCTGGGCGAG 420
QY 993 CTCACAAGCTCTCGGGAAGGCGCTCCAGCAGCGCATCTTCTCTGTTCAACACGTC 1052
Db 421 CTCACAAGCTCTCGGGAAGGCGCTTCAGCAGCGCATCTTCTCTGTTCAACAGTGC 480
QY 1053 CTGCTATACAGCGCGGCGCTGACGGCTCCAATCAGTTTAAAGTCCACGGGAGCTC 1112
Db 481 TTGCTGTATACAGCGCGGCGCTGACAGCATCTAATCAGTTTAAAGTCCACGGAGCTC 540
QY 1113 CGGCTCTATGCGATCAGCATTTGAGGAGCGGAGAGCA 1150
Db 541 CCACCTCTATGCGATCAGCATGAGGAGAGTGGAGGGA 578

RESULT 4
BC009153
LOCUS BC009153 3719 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304
IMAGE:2655209, mRNA, complete cds.
ACCESSION BC009153
VERSION BC009153.1 GI:14318718
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3719)
Strausberg, R.
Direct Submission
Submitted (05-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 5 Row: m Column: 20.

FEATURES

Location/Qualifiers

1..3719

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="FVB/N-3"

/clone="MGC:6304 IMAGE:2655209"

/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."

/clone_lib="NCI CGAP_Mam2"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

99..3296

CDS

/codon_start=1

/product="Similar to KIAA0793 gene product"

/protein_id="AAH09153.1"

/db_xref="GI:14318719"

/translation="MGEIEGTYRALPTSGRLGGQTAIGVSTLEPQSLSPRMQEKHM
RIRVLDSVLEFIEPKDGOVLLTQWKKHLNTECDYFGLKFNKVSQYIWLPEPM
KPIIRQVRKPNVLRVAVKPPDPGLOOEYTRYLPALQRLLERLTCTANTA
ALLSHLQSEIGDYEDTLDRHLKANEVLPNOEKSLKILDFHQRHTGOTPAESDFQ
VLEIARKLEMGIRHMSADREGTKINLAVSHMGVLVFGQTKINTFNKSKVRKLSFR
RKRLKLHPEVHGYPQDTLEFLLSRDECKNFKWICVEYHFFRLSDQPKPAKAVE
FSGSFRHYRQKQVLDVYKDGGMKPIYERHRSKTRTSLHALTDLPKQSVSFTD
GLRTSASISANVSYPSPSSLPGLPNLKDSSSLVDPOAPVTKSTAARSQGPS
SSDGSTQSALHPGPVLPVPGFSDPSPSPSLKSHLSLCPLOALALSTAEQGS
PVLSPVLSAGTARMNDQEEOKHMPDEAYEIAKEILATERVYLKOLEVTVWFRS
VLJKEAPMALMALLFSNIDPVEFHRGFLHEVEQRALWEGSPSAHLKGHQIRIGD
ILRNMRHQKFTSFORHDEVLTELEKATKCKKLEAVYEFELQKVCYLPENLFL
KPVRLHYRLLSLCAHYSPGHRDYADCHALKAITVETVTELOQSLRLENLQKLT
EQRDLVGENIAPGREFIREGKHLTKKGLQRMFFLFSDMLLYTSKYSTGASHF
RIRGLPLRGLMVESENEWSVLCFTIYAAQKTIIVAASTRLEKEKQMODLNAIQ
ACTIGDSPVLLGPGVYRTPRSSDEVSEEDSGRNGRSELEGNHRAHNTMHVCW
YRNTSVSRADHAAVENOLSGYLLRKFKNNGWOKLVWVETNFCLEFFVYKTHODDYLA
SLPLGLYSVSLPREADSHKDYFKLQPKSHVYFFRAESKYTFERMMOVIKRASSPG
RPPSTQCSHSPGLEAIREKEACPPSLDKNL"

BASE COUNT 997 a 979 c 929 g 814 t

ORIGIN

Query Match

Best Local Similarity 29.2%; Score 335.6; DB 10; Length 3719;

Matches 576; Conservative 0; Mismatches 379; Indels 3; Gaps 1;

QY 196 CCAACCTGAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCGCGCTGCTGAATGACC 255
Db 1591 CCGCGCTGAGTACAGTGTAGCAGGGTGCATCCCTCTGCTCAGCCCTGTCTCAGTGGCG 1650
QY 256 AGGCTGCCCCCGAGCGAGGATGAGTGGCGCGGAGGAAGATGCCCAACTGATA 315
Db 1651 CTGGTACAGCGAGGATGGACACCAAGAAGAGCAGAGAACACACACATGCCCAAGATG 1710
QY 316 AAGCGTACTTCATAGTAAAGAGTCTCTACCAAGCGGCGAAATATCTCAAGGATCTCG 375
Db 1711 AGCCCTATTTTCATAGCAAGAGGATCTCGCTACAGCAAGCAAGCTTGAAGGATTTAG 1770
QY 376 AAGTTATCACTTCCTGTTTCAGACACACAGTGTAGCAAGAGGAGCGCATCGCCGGAAGCAC 435
Db 1771 AAGTTATCACTGTGTTCCGAGCGCTGTGTATCAAGGAGGAGCGCATCGCTGCAGCCC 1830
QY 436 TGAAGAGTCTCATATTTCCCGAATTTTGAACCTTTGGACAAATTTTCACTAATTTTCTCA 495
Db 1831 TGATGGCCCTGCTTTCTCCAACTGATCCAGTCTACGAGTTCCACAGAGGCTTCTCTTC 1890
QY 496 AGGAATTTGAGCAACACACTTGCCTGTGGAGGCGGCTCAATGCCCCA---AATCAGAG 552
Db 1891 ACAGGTGGAACAGAGGCTGGCACTCTGGGAGGCGGCTCCAGTGCCCACTTAAAGAGTG 1950

QY 553 ATTACCAAGAATCGGCGATGTCTATGCTGAAGAACAATTCAGGGCATGAAGCACCTTGGCGG 612
Db 1951 ATCACCAGCAATCGGGACATCTCTCTAGGAACATCGCTAGTTAAAGGAATTTACTA 2010
QY 613 CTCACCTGTGGAACACAGACGAGGCGCTTGGAGCCCTTGGAGAAATGAATCAAGAGTCTCC 672
Db 2011 GCTACTTCCAAAGACACGATGAGGTCTCTAACAGAACTTGGAAAAGGCCACAAAACACTGA 2070
QY 673 GCGCGGTGGAGAATTCTCTCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCTCCGCTCA 732
Db 2071 AAAAGCTGGAGGAGCTTACAAAGAGTTTGAGCTCCAAAAGGTCTGCTACTCTCCCTCTCA 2130
QY 733 ACACCTTCTCTCTCGCGGCGCTCATCGCGCTCATGCACTACAAGCAGGTCCTCGAGCGGC 792
Db 2131 ACACATTTCTGTCTGAAGCCGCTCCAGAGGCTAGTCCACTACCGCTCTGCTGCTGAGCGGC 2190
QY 793 TGTGCAACACACCACCGCGAGCCAGCGACTTCAGGACTCCCGAGCGGCTTTTGGCAG 852
Db 2191 TGTGTGCTCACTACTCTCTCTGGCAGCGGACTATGCGGACTGCCATGAGGCACATGAAG 2250
QY 853 AGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGA 912
Db 2251 CCATCAGAGAAGTGACCCAGCGCTCCAGCAAGCCCTTACCCGCGCTGGAAAACCTACAGA 2310
QY 913 AGCTGCACGAACCTCAAGAAAGATTGTGATGGCATTGACAAATCTTGTGGTTCCGGGAAGGG 972
Db 2311 AATTGAGGAGCTACAGCGAGACCTGTCGGTGTAGAAAACCTCATGCTCTCTGGGAGGG 2370
QY 973 AGTTCATCGTCTGGGCGACCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGT 1032
Db 2371 AGTTTATCGTGGAGGCTGCTGCACAAAGCTCACCAAGAGGCGCTGCAACACAGGATGT 2430
QY 1033 TCTTCTGTTTCACAGCTCTCTCTATACAGAGCGCGGGGCTGACGGGCTTCCAATCAGT 1092
Db 2431 TTTTCTGTTCTCAGATATGTTTACTGTATACAAGAAAAGTGTACAGAGGACCGACGTCATT 2490
QY 1093 TTAAGTCCACGGGAGCTCCCTCTCTATGGCATGACGATTGAGGAGAGGAGGAAGAGA 1150
Db 2491 TCCGGATCGTGGCTTCCTTCCACTCCGTCGATGCTGTAGAGAAAGTGTAGGAATGA 2548
RESULT 5
AB018336
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
FEATURES
source
AB018336
Homo sapiens mRNA for KIAA0793 protein, complete cds.
AB018336
AB018336.1 GI:3882306
Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hk05692.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kohani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
DNA Res. 5 (5), 277-286 (1998)
2 (bases 1 to 3997)
Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
Direct Submission
Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
Location/Qualifiers
1..3997
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT	998 a	1132 c	1068 g	799 t
------------	-------	--------	--------	-------

Gaps 2

SOURCE ORGANISM

REFERENCE

REFERENCE AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

Db	2087	TGGAGGCAGTGTACAGGAGTTT	GAGCTGCAGAGGTTCTGCTACTTCTGCTCTCAACACGT	2146
Qy	739	TCCTCCTCGCGCCACTGCACCGGCTCATATGCACTACAAGCAGGTCCTTGGAGCGGCTGTGCA	798	
Db	2147	TCCTGCTGAAGCCATCCAGGGCTGCTGCACCTACCGCTTCTGCTGCGCGGCTATGCG	2206	
Qy	799	AACACACCCGCGGAGCCAGCGGACTTCAGGCACTGCCAGCGCGCTTTGGCAGAGATCA	858	
Db	2207	GACATTACAGCCCGGCGCACCATGACTACGTGACTGCCATGACGCCCTCAAGGCCATCA	2266	
Qy	859	CGGAGATGGTGGCACACGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCACAGAGCTGC	918	
Db	2267	CAGAGTGCACCCACACACTACAGCACATTCATCCCGGCTGGAGAACCTCGAAGGATAA	2326	
Qy	919	ACGAACCTCAAGAAAGATTTGATTGGCAATTCACAACTCTGTGGTTCGCGGAGGAGGATTCA	978	
Db	2327	CGGAGCTGCAGCGGCACTGGTGGGCATAGAGAACCTCATTTGCTCTGGCAGGAGTTCA	2386	
Qy	979	TCCGCTCTGGCAGCGCTCAGCAAGCTCTCGGGGAGGCGGCTCCAGCAGCGCATTTCTTCC	1038	
Db	2387	TCCGTGAGGGTGCTTTCACAAAGCTACCAAGAAGGCGCTGCAGCAGAGCATCTTTTTC	2446	
Qy	1039	TGTTCAAGCAGGCTCCTGCTATACAGACCGCGGGGCTGACGGGCTCCCAATCAGTTTAAAG	1098	
Db	2447	TGTTCTCAGATATGTTGCTGTACACAGAAAGAGGATGTCAGGAGCACGACCTCCGGA	2506	
Qy	1099	TCCACGGGCACTCCCGCTCTATGCGATGACGATTGAGGAGAGCGAAGACGA	1150	
Db	2507	TCCGGGCGCTTCCTCCCGCTCCAGGCACTGCTGGTGGAGAAAGTGATACGA	2558	
RESULT 6				
LOCUS	BC027077			
DEFINITION	Mus musculus, clone IMAGE:5010882, mRNA, partial cds.			
ACCESSION	BC027077			
VERSION	BC027077.1	GI:20071584		
KEYWORDS				
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
TITLE	1 (bases 1 to 2431)			
JOURNAL	Direct Submission			
	Submitted (04-APR-2002) National Institutes of Health, Mammalian			
	Gene Collection (MGC), Cancer Genomics Office, National Cancer			
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
	USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: Gilbert Smith, Ph.D.			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Baylor College of Medicine Human Genome			
	Sequencing Center			
	Center code: BCM-HGSC			
	Web site: http://www.hgsc.bcm.tmc.edu/cdna/			
	Contact: amg@bcm.tmc.edu			
	Gunarathne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,			
	Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,			
	Richards, S., Gibbs, R.A.			
	Clone distribution: MGC clone distribution information can be found			
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	Series: IRAC Plate: 45 Row: c Column: 12			
	This clone was selected for full length sequencing because it			
	passed the following selection criteria: Hexamer frequency ORF			
	analysis.			
FEATURES	Location/Qualifiers			
source	1..2431			
	/organism="Mus musculus"			

```
/db_xref="taxon:10090"
/map="FVB/N-3"
/clone="IMAGE:5010682"
/tissue_type="mammary tumor. MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<!.1065
CDS
/codon_start=1
/product="Unknown (protein for IMAGE:5010682)"
/protein_id="AAH27077.1"
/db_xref="GI:20071585"
/translation="YKHLERLCKHPPNHADFRCAALAEITEMVAQLHGTMIKME
NFQKLEKLDIGDNLVTPGRFELKLSKGLQORMFFLNDVLLYTSRGL
TASNOFKHGLPLXGMITEESEEWGPHCLTLRGQSIIVAASSRSRSEKMWEDI
QMAIDLAKNSGTPPELLASSPDNKSDEATAADQSEDDLSASRTSLERQAPHRGN
TWVHCWHRSTSVSWDESIAVENQLSKNLLRFRNKGWKLWVFNFCILFFYKSH
QDSHPLASLPGLSLTTPSESENHDKYVFLKFKSHVYFRAESEYTFERWMEVIR
SATSSASRAHILSHRESHLY"
BASE COUNT      581 a   659 c   614 g   577 t
ORIGIN
Query Match      27.2%; Score 312.8; DB 10; Length 2431;
Best Local Similarity 88.9%; Pred. No. 1.5e-64;
Matches 338; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 771 TACAAGCAGGCTCTGGAGCGGCTGTGCAACACCACCCCGGAGCCAGCCGACTTCAGG 830
Db 1 TATAAGCATGTCTGGAGAGGCTGTGCAAGCACACCCACCAACCCAGCCGACTTCAGG 60
QY 831 GACTGCGGAGCCGCTTTGGCAGAGATCAGGAGATGGTGCGACAGCTCCACCGGTACGATG 890
Db 61 GACTGCGAGCTCGCTGGCGGAGATCAGAGATGGTGCGACAGCTGATGATACCATG 120
QY 891 ATCAAGATGAGAAATTTCCAGAACTGCGACGAACTCAAGAAATTTGATGGCATTTGAC 950
Db 121 ATCAAGATGAGAACTTCCAGAACTGCGATGAGCTCAAGAAATTTGATGGCATTTGAC 180
QY 951 AATCTTTGTTGCTCCGGAAGGAGTTTCATCCGTCGGCAGCCTCAGCAAGCTCTCGGG 1010
Db 181 AATCTTTGATCCCGAGGAGGAGTTTCATCCGCTGGGAGCCTCAGCAAGCTCTCGGG 240
QY 1011 AAGGGCTCAGCAGCGCATGTTCTTCCTGTTCAACGAGCTCCTGCTATACAGAGCCGG 1070
Db 241 AAGGGCTCAGCAGCGCATGTTCTTCCTGTTCAACGAGCTCCTGCTGTTATACAGCCGG 300
QY 1071 GGCTGACGCGCTCCATCAGTTTAAGTCCAGGCGCAGCTCCGCTCTATGGCATGACG 1130
Db 301 GGACTGACGCGCATCTAATCAGTTTAAAGTCCAGGAGAGCTCCCACTCTATGGCATGACG 360
QY 1131 ATTGAGGAGCGGAAGACGA 1150
Db 361 ATCAGGAGAGTACGAGGA 380
RESULT 7
AL137249
LOCUS      106578 bp      DNA      linear      PRI 12-FEB-2002
DEFINITION Human DNA sequence from clone RP11-111L24 on chromosome
13q31.3-32.3, complete sequence.
ACCESSION  AL137249
VERSION     AL137249.29 GI:18655955
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 106578)
AUTHORS     Garner,P.
TITLE       Direct Submission
JOURNAL     Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
```

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced gi:18121451.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-111L24 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-111L24. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-295B17 is at 104579 in this
sequence. The true right end of clone RP11-261P24 is at 2000 in
this sequence.

FEATURES

Location/Qualifiers
1..106578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q31.3-32.3"
/clone="RP11-111L24"
/clone_lib="RPCI-11.1"
79533
misc_feature
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 800bp by restriction digest data"

BASE COUNT 28669 a 25979 c 25140 g 26790 t
ORIGIN
Query Match 21.0%; Score 241; DB 9; Length 106578;
Best Local Similarity 100.0%; Pred. No. 3.5e-47;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 AAGCACCCTGGCGGCTCACCTGTGGAAGCAGCAGCGGCGCTTGGAGCCCTCGAGATGGA 659
|||||
Db 37298 AAGCACCCTGGCGGCTCACCTGTGGAAGCAGCAGCGGCGCTTGGAGCCCTCGAGATGGA 37357
|||||
QY 660 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGGTGTGT 719
|||||
Db 37358 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTTCAGCTGCAGAAAGGTGTGT 37417
|||||
QY 720 TACCTACCGCTCAACACCTTTCCTTCCTGCGGCCACTGCACCGGCTCATGCACTACAAAGCAG 779
|||||
Db 37418 TACCTACCGCTCAACACCTTTCCTTCCTGCGGCCACTGCACCGGCTCATGCACTACAAAGCAG 37477
|||||
QY 780 GTCCTGGAGCGGTGTGGAACACACACCGCGGAGCCGCGGCTTCAGGGACTCCCGCA 839
|||||
Db 37478 GTCCTGGAGCGGTGTGGAACACACACCGCGGAGCCGCGGCTTCAGGGACTCCCGCA 37537
|||||
QY 840 G 840
Db 37538 G 37538

RESULT 8

AC094777

LOCUS

DEFINITION

AC094777 224310 bp DNA linear HTG 24-AUG-2002
 Rattus norvegicus clone CH230-4N8, *** SEQUENCING IN PROGRESS ***,
 65 unordered pieces.

ACCESSION

AC094777.6 GI:22164997

VERSION

HTG: HTGS_PHASE1.

KEYWORDS

Norway rat.

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 224310)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Duigan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G.,
 Olarinpangoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tinney, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wluczky, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, D., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,
 Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 224310)

Worley, K. C.

REFERENCE

AUTHORS

TITLE

Direct Submission

JOURNAL

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

3 (bases 1 to 224310)

Rat Genome Sequencing Consortium.

Direct Submission

JOURNAL

COMMENT

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 9, 2002 this sequence version replaced gi:22094284.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBIZ

Center clone name: CH230-4N8

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 172428 bases at least Q40

Consensus quality: 179331 bases at least Q30

Consensus quality: 184096 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1011: contig of 1011 bp in length
 * 1012 1111: gap of unknown length
 * 1112 2458: contig of 1347 bp in length
 * 2459 2558: gap of unknown length
 * 2559 3836: contig of 1278 bp in length
 * 3837 3936: gap of unknown length
 * 3937 5467: contig of 1531 bp in length
 * 5468 5567: gap of unknown length
 * 5568 6863: contig of 1296 bp in length
 * 6864 8963: gap of unknown length
 * 8964 8013: contig of 1050 bp in length
 * 8014 8113: gap of unknown length
 * 8114 9431: contig of 1318 bp in length
 * 9432 9531: gap of unknown length
 * 9532 10970: contig of 1439 bp in length
 * 10971 11070: gap of unknown length
 * 11071 12178: contig of 1108 bp in length
 * 12179 12278: gap of unknown length
 * 12279 13604: contig of 1326 bp in length
 * 13605 13704: gap of unknown length
 * 13705 15160: contig of 1456 bp in length
 * 15161 15260: gap of unknown length
 * 15261 16520: contig of 1260 bp in length
 * 16521 16620: gap of unknown length
 * 16621 18279: contig of 1659 bp in length
 * 18280 18379: gap of unknown length
 * 18380 19562: contig of 1183 bp in length
 * 19563 19662: gap of unknown length
 * 19663 20728: contig of 1066 bp in length
 * 20729 20828: gap of unknown length
 * 20829 22059: contig of 1231 bp in length
 * 22060 22159: gap of unknown length
 * 22160 24977: contig of 2818 bp in length
 * 24978 25077: gap of unknown length
 * 25078 26432: contig of 1355 bp in length
 * 26433 26532: gap of unknown length
 * 26533 27664: contig of 1132 bp in length
 * 27665 27764: gap of unknown length
 * 27765 29834: contig of 2070 bp in length
 * 29835 29935: gap of unknown length
 * 29936 32076: contig of 2142 bp in length
 * 32077 32176: gap of unknown length
 * 32177 33953: contig of 1777 bp in length
 * 33954 34053: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Homo sapiens chromosome 13, clone RP11-96B23
 2 (bases 1 to 152053)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Brown,A., Burkett,G., Castle,A.,
 Boguslavsky,L., Bouckgalter,B., Bowan,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
 DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severly,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 152053)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A.,
 Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
 Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K.,
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., McPheeters,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R., Seaman,S.,
 Severly,P., Sounguez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
 Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 4, 2000 this sequence version replaced gi:7229803.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4938
 Center clone name: 96_B_23
 ----- Summary Statistics
 Sequencing vector: M13: M77815; 93% of reads
 Sequencing vector: Plasmid: n/a; 0.0% of reads
 7.35091718610864Chemistry: Dye-terminator Big Dye; 100% of
 reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 136451 bases at least Q40
 Consensus quality: 143612 bases at least Q30
 Consensus quality: 146365 bases at least Q20
 Insert size: 158000; agarose-
 Quality coverage: 148853; sum-of-
 Quality coverage: 3.4 in Q20 bases; agarose-
 Quality cov.

TITLE
 JOURNAL
 REFERENCE
 COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 393)
AUTHORS
Olivier, M. and Cox, D.R.
TITLE
Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL
Unpublished (2000)
COMMENT

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu
Primer A: AGAGGATGCTTATCTCTCCG
Primer B: ACGGTACGATGATCAAGATGAG
STS size: 290
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
AmpliTaq Gold Polymerase: 0.07 units/uL
Total Vol: 5 uL

Buffer:

MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Finished human sequence in NCBI. STSs designed and developed at the
Stanford Human Genome Center.

FEATURES
source

1..393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
/clone_lib="Human"
12..301

primer_bind
12..34

primer_bind
complement(279..301) 115 t

BASE COUNT
80 a 105 c 93 g 115 t

Query Match 14.1%; Score 162.4; DB 11; Length 393;
Best Local Similarity 95.9%; Pred. No. 2.2e-28;
Matches 188; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 790 GCGTGTGCAACA-CCACCCGCGACGCCGACTTCAGGAGCTCCGAGCCGCTTTG 848
|||||
Db 393 GCGTGTGCAACAACCCCGCGACGCCGACTTCAGGAGCTCCGAGCCGCTTTG 334

QY 849 GCAGAGATCACGAGATGTTGGACACAGC-TCCACGGTAGCATGATCAAGATGGAATTT 907
|||||

Db 333 GCAGAGATCACGAGATGTTGGACACAGCTTCCACGGTAGCATGATCAAGATGGAATTT 274

QY 908 CCAGAAGCTGCAGAACTCAAGAAAGATTTGATTCGATTCATTCATTCGTTGTTCCCGG 967
|||||

Db 273 CCAGAAGCTGCAGAACTCAAGAAAGATTTGATTCGATTCATTCATTCGTTGTTCCCGG 214

QY 968 AAGGGAGTTTCATCCGT 983
|||||

Db 213 AAGGGTAAGCAGCAGT 198
|||||

RESULT 12

AX261839/c

LOCUS AX261839 426 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 1490 from Patent WO0173027.
ACCESSION AX261839
VERSION AX261839.1 GI:16510806
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Meagher, M.J., Xu, J. and King, G.E.
Compositions and methods for therapy and diagnosis of colon cancer
Patent: WO 0173027-A 1490 04-OCT-2001;
CORIXA CORPORATION (US)

FEATURES

Location/Qualifiers

source

1..426

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 89 a 109 c 109 g 119 t

ORIGIN

Query Match 13.4%; Score 154.6; DB 6; Length 426;
Best Local Similarity 61.9%; Pred. No. 1.6e-26;
Matches 263; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 321 TACTTCATAGCTAAGGAAGTGTCTACCCGAGCGAGCAACATATCTGAAGCATCTCGAAGTT 380
|||||

Db 426 TACTTCATAGCTAAGGAAGTGTCTCGCTACAGAACGCAACATCTCAAGGATTTAGAAGTT 367
|||||

QY 381 ATCACTTCTGCTTTTCAGAGCACAGTGTAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAA 440
|||||

Db 366 ATTACCGTGTGTTCCGCGAGCGCAGTGGTGAAGAGGAGCGCCATGCCCTCGACTCTGATG 307
|||||

QY 441 AGTCATATATCCCGAATTTTGAACCTTTGACAAATTTTCATCTACTACTTCTCAAGGAA 500
|||||

Db 306 AGCTGCTCTTCTCCAAACATCGATCCATCTATGATTCACAGAGGCTTCCTGCGCGAG 247
|||||

QY 501 ATTGAGCAACGACTTGCCTGTGGGAAGCGCTCAATGCCCAATACAG-ATTATAC 557
|||||

Db 246 GTGGAGCAGAGCTGGCACTCTGGGAAGGCGCTCCAAAGCCCAACAAAGGAGCTAT 187
|||||

QY 558 CAAGAATCGCGGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTAC 617
|||||

Db 186 CAACGAATCGGGACATCTCTGCTCAGGAACATGCGCCAGTTAAAGGAGTTTACCAGCTAC 127
|||||

QY 618 CTGTGGAGCACAGCGAGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGCGG 677
|||||

Db 126 TTCCAAAGACATGACGAGATCTTAACAGAACTGGAAGGCTACCAAGCGTGTAAAGAG 67
|||||

QY 678 CTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGCTGTGTACTACCTCAACACC 737
|||||

Db 66 TTGGAGGAGCTGTACAGGAGTTTGAAGTTCAGCTGCAGAGGCTCTGCTACTTGCCTCAACAG 7
|||||

QY 738 TTCTCT 742

|||||

Db 6 TTCTCT 2

RESULT 13

AX261839/c

LOCUS AC130919 120911 bp DNA linear HTG 15-AUG-2002
DEFINITION Rattus norvegicus clone CH230-159A1, *** SEQUENCING IN PROGRESS
***, 63 unordered pieces.

ACCESSION

AC130919

VERSION

AC130919.1 GI:22261960

KEYWORDS

HTG; HTGS_PHASE1.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 120911)

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagm,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kwis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshehua,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,P., Mapa,P., Martin,K., Martin,R., Martinez,E., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reese,M., Richards,S., Riggs,F., Rives,C., Rookey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sigson,I., Sitter,C.D., Smajd,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.

TITLE

Unpublished

JOURNAL

2 (bases 1 to 120911)

REFERENCE

Rat Genome Sequencing Consortium.

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBGN
Center clone name: CH230-159A1
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 53051 bases at least Q40
Consensus quality: 57784 bases at least Q30
Consensus quality: 61330 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1028: contig of 1028 bp in length
1128: gap of unknown length
1129 1128: contig of 1037 bp in length
2165: gap of unknown length
2166 2165: contig of 1355 bp in length
3620: contig of 1355 bp in length
3621 3620: gap of unknown length
5093: contig of 1373 bp in length
5193: gap of unknown length
5094 5193: contig of 1387 bp in length
5194 5193: gap of unknown length
6581 5194: contig of 1489 bp in length
6681 5194: contig of 1489 bp in length
8170 8169: gap of unknown length
8270 8169: gap of unknown length
8987: contig of 1628 bp in length
8989 8987: gap of unknown length
9998 8989: contig of 1333 bp in length
11330 9998: gap of unknown length
11431 11330: contig of 1654 bp in length
13085 11431: gap of unknown length
13184 13085: contig of 1348 bp in length
14532 13184: gap of unknown length
14633 14532: gap of unknown length
15891 14633: contig of 1259 bp in length
15991 15891: gap of unknown length
17101 15991: contig of 1110 bp in length
17201 17101: gap of unknown length
18776 17201: contig of 1575 bp in length
18777 18776: gap of unknown length
20138 18777: contig of 1262 bp in length
20238 20138: gap of unknown length
21285 20238: contig of 1047 bp in length
21385 21285: gap of unknown length
26330 21385: contig of 1245 bp in length
27330 26330: gap of unknown length
27730 27330: contig of 1053 bp in length
23784 27730: gap of unknown length
25084 23784: contig of 1201 bp in length
25184 25084: gap of unknown length
26257 25184: contig of 1073 bp in length
26357 26257: gap of unknown length
27942 26357: contig of 1385 bp in length
28042 27942: gap of unknown length
29312 28042: contig of 1270 bp in length
29412 29312: gap of unknown length
30808 29412: contig of 1396 bp in length
30908 30808: gap of unknown length
32522 30908: contig of 1614 bp in length
32623 32522: gap of unknown length
34182 32623: contig of 1560 bp in length
34282 34182: gap of unknown length
36102 34282: contig of 1820 bp in length
36202 36102: gap of unknown length
37338 36202: contig of 1136 bp in length
37438 37338: gap of unknown length
38559 37438: contig of 1121 bp in length
38560 38559: gap of unknown length
40454 38560: contig of 1795 bp in length
40554 40454: gap of unknown length
42228 40554: contig of 1674 bp in length
42329 42228: gap of unknown length
43442 42329: contig of 1114 bp in length
43542 43442: gap of unknown length
45213 43542: contig of 1671 bp in length
45313 45213: gap of unknown length
47351 45313: contig of 2038 bp in length
47352 47351: gap of unknown length

Consensus quality: 186552 bases at least Q20
Insert size: 206000; agarose-fp
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 6.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 698: contig of 698 bp in length
699 798: gap of 100 bp
799 1437: contig of 639 bp in length
1438 1537: gap of 100 bp
1538 2255: contig of 718 bp in length
2256 2355: gap of 100 bp
2356 3033: contig of 678 bp in length
3034 3133: gap of 100 bp
3134 3783: contig of 650 bp in length
3784 3883: gap of 100 bp
3884 4559: contig of 676 bp in length
4560 4659: gap of 100 bp
4660 5309: contig of 650 bp in length
5310 5409: gap of 100 bp
5410 6053: contig of 644 bp in length
6054 6153: gap of 100 bp
6154 7009: contig of 856 bp in length
7010 7109: gap of 100 bp
7110 7747: contig of 638 bp in length
7748 7847: gap of 100 bp
7848 8125: contig of 278 bp in length
8126 8225: gap of 100 bp
8226 8956: contig of 731 bp in length
8957 9056: gap of 100 bp
9057 10142: contig of 1086 bp in length
10143 10242: gap of 100 bp in length
10243 11404: contig of 1162 bp in length
11405 11504: gap of 100 bp
11505 12593: contig of 1089 bp in length
12594 12693: gap of 100 bp
12694 14681: contig of 1988 bp in length
14682 14781: gap of 100 bp
14782 17534: contig of 2753 bp in length
17535 17634: gap of 100 bp
17635 21928: contig of 4294 bp in length
21929 22028: gap of 100 bp
22029 29432: contig of 7404 bp in length
29433 29532: gap of 100 bp
29533 35637: contig of 6105 bp in length
35638 35737: gap of 100 bp
35738 84581: contig of 48844 bp in length
84582 84681: gap of 100 bp
84682 93581: contig of 8900 bp in length
93582 93681: gap of 100 bp
93682 101846: contig of 8165 bp in length
101847 101946: gap of 100 bp
101947 111400: contig of 9434 bp in length
111401 111500: gap of 100 bp
111501 131247: contig of 19747 bp in length
131248 131347: gap of 100 bp
131348 151032: contig of 19685 bp in length
151033 151132: gap of 100 bp
151133 184681: contig of 33549 bp in length
184682 184781: gap of 100 bp
184782 190588: contig of 5807 bp in length.

FEATURES
Source

1. .190588
/organism="Mus musculus"
/db_xref="taxon:10090"

/clone="RP23-321M11"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .698
/note="assembly_fragment"
799. .1437
/note="assembly_fragment"
1538. .2255
/note="assembly_fragment"
2356. .3033
/note="assembly_fragment"
3134. .3783
/note="assembly_fragment"
3884. .4559
/note="assembly_fragment"
4660. .5309
/note="assembly_fragment"
5410. .6053
/note="assembly_fragment"
6154. .7009
/note="assembly_fragment"
7110. .7747
/note="assembly_fragment"
7848. .8125
/note="assembly_fragment"
8226. .8956
/note="assembly_fragment"
9057. .10142
/note="assembly_fragment"
10243. .11404
/note="assembly_fragment"
11505. .12593
/note="assembly_fragment"
12694. .14681
/note="assembly_fragment"
14782. .17534
/note="assembly_fragment"
17635. .21928
/note="assembly_fragment"
22029. .29432
/note="assembly_fragment"
29533. .35637
/note="assembly_fragment"
35738. .84581
/note="assembly_fragment"
84682. .93581
/note="assembly_fragment"
93682. .101846

Query Match 12.9%; Score 148.6; DB 2; Length 190588;
Best Local Similarity 83.3%; Pred. No. 6e-25;
Matches 169; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 97 CGCAGCCACGACAGGCTCCCTGACTGGCAGTCTCTACCTTTCCGAGCTCTGTGNACT 156
Db 187607 CTATCCCCAACACAGGCTCCCTGACTGGTAGCCCTACCTTTTCAGAGCTCTCATCACT 187548
QY 157 CGCAGGGGGAGTGGCCCTGCCAACGTGACCTTGCTCCCAACCTTGAGCCCGCACCA 216
Db 187547 CACAGGAGGAGGCGCGCCAGCAGCAACGTGACCTGTCTCCCACTGAGTCCTGACACA 187488
QY 217 AGCAGGCTCTCCCTTGTATCAGCCCGCTGCTGAATGACCAAGGCTGCCCGGACGACG 276
Db 187487 AGCAGGCTCTCCCTGTATCAGCCCGCTCTCTGAACGACCAAGCGTGCCCAAGGACGACG 187428
QY 277 ATGAGATGAGGCGCGGAGGAAG 299
Db 187427 ACGAAGAAGAGGCGCGGGAAG 187405

RESULT 15
AC122885/c 221502 bp DNA linear HTG 23-JUN-2002
LOCUS AC122885 221502 bp DNA linear HTG 23-JUN-2002
DEFINITION Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT

SEQUENCE, 8 unordered pieces.
AC122885
VERSION
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 221502)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 221502)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 221502)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jun 23, 2002 this sequence version replaced gi:21218511.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0090H14
----- Summary Statistics -----
Sequencing vector: W13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 216976 bases at least Q40
Consensus quality: 217754 bases at least Q30
Consensus quality: 218341 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 221868; sum-of-contigs
Quality coverage: 15.60 in Q20 bases; agarose-fp
Quality coverage: 12.41 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8441: contig of 8441 bp in length
* 8442 8541: gap of unknown length
* 8542 27277: contig of 18736 bp in length
* 27278 27377: gap of unknown length
* 27378 42339: contig of 14962 bp in length
* 42340 42439: gap of unknown length
* 42440 71538: contig of 29099 bp in length
* 71539 71638: gap of unknown length
* 71639 99587: contig of 27949 bp in length
* 99588 99687: gap of unknown length
* 99688 134048: contig of 34361 bp in length
* 134049 134148: gap of unknown length
* 134149 221349: contig of 87201 bp in length
* 221350 221449: gap of unknown length
* 221450 221502: contig of 53 bp in length.

FEATURES
Source

1..221502
/organism="Mus musculus"
/db_xref="taxon:10090"

/chromosome="UNK"
/clone="RP23-90H14"
1..8441
/note="assembly_name:Contig59"
8542..27277
/note="assembly_name:Contig60"
27378..42339
/note="assembly_name:Contig61"
42440..71538
/note="assembly_name:Contig62"
71639..99587
/note="assembly_name:Contig63"
99688..134048
/note="assembly_name:Contig64"
134149..221349
/note="assembly_name:Contig65"
221450..221502
/note="assembly_name:Contig19"
752 others
BASE COUNT 61405 a 54522 c 50450 g 54373 t
ORIGIN
Query Match 12.9%; Score 148.6; DB 2; Length 221502;
Best Local Similarity 83.3%; Pred. No. 6.1e-25;
Matches 169; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 97 CGCAGCCAAAGCAGGCTCCCTGACTGGCAGTCCTCACCTTCCGAGCTGTCTGTGAAC 156
Db 113505 CTATCCCCCAAGAGGCTCCCTGACTGGTAGCCCTCACCTTTTCAGAGCTGTCCATCAAC 113446
QY 157 CGCAGGGGGAGTGGCCCTGCGCAAGTGTCTCCCAACCTGAGCCCGGACCA 216
Db 113445 CACAGGAGGAGCGCCGCCAGCAAGTGACCTGTCTCCCAACCTGAGTCTCTGACAACA 113386
QY 217 AGCAGGCTCTCTGTGATCAGCCCTGCTGAATGACAGGCTGCCCGGACGAGC 276
Db 113385 AGCAGGCTCTCTGTGATCAGCCCTGCTGAATGACAGGCTGCCCGGACGAGC 113326
QY 277 ATGAGGATGAGGCGCGGAGGAG 299
Db 113325 ACGAAGAGAGGCGCGGAGGAG 113303
Search completed: December 3, 2002, 21:05:54
Job time : 3209.8 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 14:48:04 ; Search time 181.787 Seconds
(without alignments)
14246.337 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1352_2501

Perfect score: 1150

Sequence: 1 cgggtacaagcaggcgac.....attgaggagagcgaagacga 1150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1150	100.0	3442	20	AAAX79183 Human chondrocyte-
2	1150	100.0	4687	21	AAAO8582 Human cytoskeleton
3	1146.8	99.7	3187	23	AAAS64760 DNA encoding novel
4	762	66.3	3094	21	AAC98992 Human pancreatic c
c 5	154.6	13.4	426	23	AAAS58814 cDNA #1490 encodin
c 6	112	9.7	1824	23	AAAS4761 DNA encoding novel
7	106	9.2	1718	22	AAAS27016 cDNA encoding nove
8	105.8	9.2	1747	23	ABLO3867 Drosophila melanog
c 9	105.8	9.2	4150	23	ABLO3866 Drosophila melanog

10	105.2	9.1	716	22	AAS27440 cDNA encoding nove
11	98	8.5	717	24	ABL90828 Human polynucleoti
12	90.2	7.8	544	22	AAH98023 Murine 7-transmemb
13	83.6	7.3	665	22	AAH07867 Human cDNA clone (
14	83.6	7.3	1296	24	ABK49898 Human cDNA encodin
15	83.6	7.3	1821	22	AAH15014 Human cDNA sequenc
16	83.6	7.3	3610	24	ABN59933 Novel human coding
c 17	72.8	6.3	462	24	ABK53702 Human eosinophil-m
18	64.4	5.6	2686	22	AAAS27013 cDNA encoding nove
19	64.4	5.6	3321	22	AAAS44718 Human full-length
20	64	5.6	1091	22	AAAS27281 cDNA encoding nove
21	60.8	5.3	1860	22	AAF99965 Human colon carcin
22	60.8	5.3	5923	22	AAAS02052 DNA encoding molec
c 23	60.8	5.3	5923	22	AAAS02052 DNA encoding molec
24	60	5.2	60	24	ABN41498 Human spliced tran
25	56.2	4.9	365	14	AAQ39923 Expressed Sequence
26	56.2	4.9	365	14	AAQ59335 Human brain Expres
27	53.2	4.6	417	22	ABA08384 Human collybistin
c 28	46.4	4.0	454	23	ABV54945 Human prostate exp
29	46.4	4.0	541	24	ABN64838 Human cancer relat
30	44.4	3.9	342	21	AAF08796 Fusarium venenatum
31	44	3.8	1609	23	ABL25953 Drosophila melanog
32	44	3.8	3381	23	ABL25950 Drosophila melanog
33	44	3.8	3609	23	ABL25952 Drosophila melanog
34	43.4	3.8	3786	23	AAAS4228 Pseudomonas aerugi
c 35	42.4	3.7	1730	22	AAAS27017 cDNA encoding nove
36	42.4	3.7	1730	22	AAAS27441 cDNA encoding nove
37	42.4	3.7	1730	22	AAI62819 Human cDNA SEQ ID
c 38	42.4	3.7	1778	22	AAI60338 Human polynucleoti
39	42.4	3.7	1816	22	AAI58552 Human polynucleoti
40	42	3.7	3306	23	ABL12855 Drosophila melanog
41	41.2	3.6	2159	15	AAQ67223 Mouse p53Nuc. Mus
42	40.8	3.5	625	22	AAH07308 Human cDNA clone (
43	40.8	3.5	1444	22	AAH41863 Human chondroadher
44	40.8	3.5	1696	22	AAI93904 Human stomach canc
45	40.8	3.5	1696	22	AAH16459 Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAAX79183
ID AAAX79183 standard; DNA; 3442 BP.

XX

AC AAAX79183;

XX

DT 17-AUG-1999 (first entry)

XX

XX Human chondrocyte-derived gene CDEP.

DE

DE Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;

XX

KW Dbl homology domain; pleckstrin homology domain; rheumatoid arthritis;

KW drug; ss.

XX

OS Homo sapiens.

XX

PN WO9928458-A1.

XX

PD 10-JUN-1999.

XX

PF 27-NOV-1998; 98WO-JP05348.

XX

PR 27-NOV-1997; 97JP-0342060.

XX

PA (CHUS) CHUGAI SEIYAKU KK.

XX

PI Kato Y, Kawamoto T, Koyano Y;

XX

WPI: 1999-371117/31.

DR

DR P-PSDB; AA07482.

XX

PT Protein CDEP expressed in differentiated chondrocytes, and gene

PT encoding it
XX Claim 5; Fig 1; 59pp; Japanese.
XX This sequence represents the coding region for a protein (CDEP) expressed
CC in differentiated human fetal chondrocytes, which contains an ezrin-like
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.
CC The nucleic acid or protein can be used in the investigation and
CC treatment of cancers and arthritic diseases (including chronic rheumatoid
CC arthritis), or for screening of candidate anticancer drugs.
XX Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;
SQ Query Match 100.0%; Score 1150; DB 20; Length 3442;
Best Local Similarity 100.0%; Pred. No. 7.1e-307;
Matches 1150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGGTAAACAGCAGCGGAGCGGCGCTCGCGCCCAACGAGGAGGAGGTCG 60
DB 1352 CGGGTAAACAGCAGCGGAGCGGCGCTCGCGCCCAACGAGGAGGAGGTCG 1411
QY 61 TTAAGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGCAGCAGGCTCCCTGA 120
DB 1412 TTAAGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGCAGCAGGCTCCCTGA 1471
QY 121 CTGGCAGTCTCACCTTTCGAGCTGTCTGTGAACCTCGCAGGGGAGTGGCCCTGCCA 180
DB 1472 CTGGCAGTCTCACCTTTCGAGCTGTCTGTGAACCTCGCAGGGGAGTGGCCCTGCCA 1531
QY 181 AGTGACCTTGTCTCCCAACCTAGCGCCCGCAGCAGGAGGAGGAGGTCG 240
DB 1532 AGTGACCTTGTCTCCCAACCTAGCGCCCGCAGCAGGAGGAGGAGGTCG 1591
QY 241 CGCTGCTGAATGACGAGGCTGCGCCCGCAGCAGGAGGAGGAGGAGGTCG 300
DB 1592 CGCTGCTGAATGACGAGGCTGCGCCCGCAGCAGGAGGAGGAGGAGGTCG 1651
QY 301 GATTCCTCAACTGATAAGCGTACTTCTATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 360
DB 1652 GATTCCTCAACTGATAAGCGTACTTCTATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 1711
QY 361 ATCTGAAGGATCTCGAAGTATACCTTCTGTTTTCAGAGCAGTGACGAAGAGGAG 420
DB 1712 ATCTGAAGGATCTCGAAGTATACCTTCTGTTTTCAGAGCAGTGACGAAGAGGAG 1771
QY 421 CCATGCGGAGACACTGAAAGTCTCATATTCGCGAATTTTGAACCTTTGCACAAATTC 480
DB 1772 CCATGCGGAGACACTGAAAGTCTCATATTCGCGAATTTTGAACCTTTGCACAAATTC 1831
QY 481 ATACTAATTTTCTAAGGAAATGAGCAACGACTTGCCTGTGGGAAGCGCGCTCAAAATG 540
DB 1832 ATACTAATTTTCTAAGGAAATGAGCAACGACTTGCCTGTGGGAAGCGCGCTCAAAATG 1891
QY 541 CCCAATCAGAGATTACCAAGAATGCGGATGTCTATGCTGAAGAACATTCAGGGCATGA 600
DB 1892 CCCAATCAGAGATTACCAAGAATGCGGATGTCTATGCTGAAGAACATTCAGGGCATGA 1951
QY 601 AGCAGCTGGCGGCTACCTGTGSAAGCAGCAGGAGGCTTGGAGCCCTGGAGATGGAA 660
DB 1952 AGCAGCTGGCGGCTACCTGTGSAAGCAGCAGGAGGCTTGGAGCCCTGGAGATGGAA 2011
QY 661 TCAAGAGCTCCGCGGCTGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAGGCTGTGT 720
DB 2012 TCAAGAGCTCCGCGGCTGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAGGCTGTGT 2071
QY 721 ACCTACCGCTCAACAGCTTCTCTCGCGGCTACGTGACCGGCTCAATGCAACAGCAG 780
DB 2072 ACCTACCGCTCAACAGCTTCTCTCGCGGCTACGTGACCGGCTCAATGCAACAGCAG 2131
QY 781 TCCTGAGCGGCTGTGCAACACCCCGCGAGCCAGCCGACTTCAGGGACTGCCAG 840
DB 2132 TCCTGAGCGGCTGTGCAACACCCCGCGAGCCAGCCGACTTCAGGGACTGCCAG 2191

QY 841 CCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGCTACGATGATCAAGATGG 900
DB 2192 CCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGCTACGATGATCAAGATGG 2251
QY 901 AGAATTTCCAGAGTGCACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 960
DB 2252 AGAATTTCCAGAGTGCACGAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 2311
QY 961 TTCGGGAAGGAGGATTCATCCGCTCTGGGAGGCTCAGCAAGCTCTCGGGAAGGGGCTCC 1020
DB 2312 TTCGGGAAGGAGGATTCATCCGCTCTGGGAGGCTCAGCAAGCTCTCGGGAAGGGGCTCC 2371
QY 1021 AGCAGCGATGTTCTTCTTCAACGAGCTCTGCTATACAGAGCCCGGGGCTGACGG 1080
DB 2372 AGCAGCGATGTTCTTCTTCAACGAGCTCTGCTATACAGAGCCCGGGGCTGACGG 2431
QY 1081 COTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGAGATTGAGGAGA 1140
DB 2432 COTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGAGATTGAGGAGA 2491
QY 1141 GCGAAGACGA 1150
DB 2492 GCGAAGACGA 2501
RESULT 2
AAA08582
ID AAA08582 standard; DNA; 4687 BP.
XX AC AAA08582;
XX AC
XX DT 19-JUL-2000 (first entry)
XX DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.
XX KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 90..3227
FT /*tag= a
FT /product= CYSKP-2
XX PN WO200017355-A2.
XX PD 30-MAR-2000.
XX PF 17-SEP-1999; 99WO-US21565.
XX PR 18-SEP-1999; 98US-0172226.
XX PR 27-APR-1999; 99US-0131321.
XX PA (INCYTE) INCYTE PHARM INC.
XX PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
XX WPI; 2000-283582/24.
XX P-PSDB; AA91947.
XX Human cytoskeleton associated proteins, used to treat cell
XX proliferative, autoimmune/inflammatory, vesicle trafficking,
XX neurological, cell motility, reproductive and muscle disorders
XX Claim 9; Page 101-102; 113pp; English.
XX AA08581-96 encode human cytoskeleton associated proteins 1 to 16
XX (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat
XX and diagnose cancer and cell proliferative, autoimmune/inflammatory,
XX vesicle trafficking, neurological, cardiovascular, cell motility,

CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CYSPK-2 to CYSPK-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CYSPK (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
CC cancers, autoimmune/anti-inflammatory disorders such as allergies, anemia,
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CYSPK antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CYSPK
CC (claimed).

XX
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;
Query Match 100.0%; Score 1150; DB 21; Length 4687;
Best Local Similarity 100.0%; Pred. No. 8.2e-307;
Matches 1150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGTAACAGCAGGGGAGCGGCGCTCGGCGCCACGAGGAAGAGGAGGTGG 60
DB 1393 CGGGTAACAGCAGGGGAGCGGCGCTCGGCGCCACGAGGAAGAGGAGGTGG 1452
QY 61 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 120
DB 1453 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 1512
QY 121 CTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGGGGAGTGGCCCTGCCA 180
DB 1513 CTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGGGGAGTGGCCCTGCCA 1572
QY 181 AGTGACCTTGCTCCCAACTGAGCCCGGACCAAGCAGGCGCTCTCCCTTGATCAGCC 240
DB 1573 AGTGACCTTGCTCCCAACTGAGCCCGGACCAAGCAGGCGCTCTCCCTTGATCAGCC 1632
QY 241 CGCTGCTGAATGACAGGCTGCGCCCGGACGAGCAGTGAAGGAGGCGGAGGAAGA 300
DB 1633 CGCTGCTGAATGACAGGCTGCGCCCGGACGAGCAGTGAAGGAGGCGGAGGAAGA 1692
QY 301 GATTCCTCAACTGATAAAGGCTACTTCATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 360
DB 1693 GATTCCTCAACTGATAAAGGCTACTTCATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 1752
QY 361 ATCTGAAGATCTGAGAGTATCATCTGCTGTTTCAGAGCAGTGAAGAGGAGGAG 420
DB 1753 ATCTGAAGATCTGAGAGTATCATCTGCTGTTTCAGAGCAGTGAAGAGGAGGAG 1812
QY 421 CCATGCGGAGCAGCTGAAAGCTCTCATATCCCGAATTTTGAACCTTTGCACAAATTC 480
DB 1813 CCATGCGGAGCAGCTGAAAGCTCTCATATCCCGAATTTTGAACCTTTGCACAAATTC 1872
QY 481 ATACTAATTTTCTAAGGAAATTTGAGCAACGACTTGCCCTGTGGGAAGCGCTCAAAATG 540
DB 1873 ATACTAATTTTCTAAGGAAATTTGAGCAACGACTTGCCCTGTGGGAAGCGCTCAAAATG 1932
QY 541 CCCAATCAGAGATTACCAAGAAATCGGGGATGTGATGTGAAGAACATTCAGGGCATGA 600
DB 1933 CCCAATCAGAGATTACCAAGAAATCGGGGATGTGATGTGAAGAACATTCAGGGCATGA 1992
QY 601 AGCACCTGCGGCTACCTGTGSAAGCAGCAGCGGCTTGGAGGCGCTGGAGATGGAA 660
DB 1993 AGCACCTGCGGCTACCTGTGSAAGCAGCAGCGGCTTGGAGGCGCTGGAGATGGAA 2052
QY 661 TCAAGAGCTCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGAGGTGTGT 720
DB 2053 TCAAGAGCTCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGAGGTGTGT 2112
QY 721 ACCTACCGCTCAACACTTCTCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGG 780
DB 2113 ACCTACCGCTCAACACTTCTCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGG 2172
QY 781 TCCTGAGCGGCTGTGCAAAACACCAACCCCGCCGAGCCAGCTTCAGGGACTGCCAG 840
DB 2173 TCCTGAGCGGCTGTGCAAAACACCAACCCCGCCGAGCCAGCTTCAGGGACTGCCAG 2232

QY 841 CCGCTTTGGCAGAGATCAGCGAGATGGTGGCACAGCTCCACGATGATCAAGATGG 900
DB 2233 CCGCTTTGGCAGAGATCAGCGAGATGGTGGCACAGCTCCACGATGATCAAGATGG 2292
QY 901 AGAATTTCCAGAGCTGCAGCAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 960
DB 2293 AGAATTTCCAGAGCTGCAGCAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGG 2352
QY 961 TTCCGGGAAGGAGTTTCATCCCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAGGGCTCC 1020
DB 2353 TTCCGGGAAGGAGTTTCATCCCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAGGGCTCC 2412
QY 1021 AGCAGCGCATGTTCTTCCTGTTTCAACAGAGTCTCTGTATACAGAGCCCGGGGTGACGG 1080
DB 2413 AGCAGCGCATGTTCTTCCTGTTTCAACAGAGTCTCTGTATACAGAGCCCGGGGTGACGG 2472
QY 1081 CCTCCAATCAGTTTAAAGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 1140
DB 2473 CCTCCAATCAGTTTAAAGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 2532
QY 1141 GCGAAGACGA 1150
DB 2533 GCGAAGACGA 2542
RESULT 3
AAS64760
ID AAS64760 standard; cDNA; 3187 BP.
XX AAS64760;
XX AC
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #564.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG00573.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 564; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS61497-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 3187 BP; 810 A; 874 C; 876 G; 627 T; 0 other:
Query Match 99.7%; Score 1146.8; DB 23; Length 3187;
Best Local Similarity 99.8%; Pred. No. 5.2e-306;
Matches 1148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CGGGTAACAGGAGCGGAGCGCGCTCGCGCGCCACCGAGGAAGAGGAGGTCG 60
DB 1353 CGGGTAACAGGAGCGGAGCGCGCTCGGAGCCACCGAGGAGGAGGTCG 1412
QY 61 TTAAGATAGGACCCAGCAGAGATAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 120
DB 1413 TTAAGATAGGACCCAGCAGAGATAACCTCAGCCCGCCGAGCAAGCAGAGCTCCCTGA 1472
QY 121 CTGGCAGTCTCACCCTTCGAGCTGCTGTGAACCTCGCAGGGGAGTGCCCTGCGCA 180
DB 1473 CTGGCAGTCTCACCCTTCGAGCTGCTGTGAACCTCGCAGGGGAGTGCCCTGCGCA 1532
QY 181 AGTGACCTTGCTCCCAACCTGAGCCCGCACCAAGCAGGCGCTCTCCCTTGATCAGCC 240
DB 1533 AGTGACCTTGCTCCCAACCTGAGCCCGCACCAAGCAGGCGCTCTCCCTTGATCAGCC 1592
QY 241 CGCTGCTGAATGACAGCGCTCGCCCGGAGCAGATGAGGATGAGGCGCGGAGGAAGA 300
DB 1593 CGCTGCTGAATGACAGCGCTCGCCCGGAGCAGATGAGGATGAGGCGCGGAGGAAGA 1652
QY 301 GATTCCTCACTGATAAGGCTACTCATAGCTAAGGAGTGCTACCCAGCGGCAACAT 360
DB 1653 GATTCCTCACTGATAAGGCTACTCATAGCTAAGGAGTGCTACCCAGCGGCAACAT 1712
QY 361 ATCTGAAGGATCTCGAAGTTATCACTTCGTGTTTCAGAGCAGTGAGCAAGAGGAGC 420
DB 1713 ATCTGAAGGATCTCGAAGTTATCACTTCGTGTTTCAGAGCAGTGAGCAAGAGGAGC 1772
QY 421 CCATGCGGAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCAACAATTC 480
DB 1773 CCATGCGGAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCAACAATTC 1832
QY 481 ATACTAATTTTCTCAAGGAAATGAGCAAGCACTTGCCCTGTGGGAAGCGCTCAAAATG 540
DB 1833 ATACTAATTTTCTCAAGGAAATGAGCAAGCACTTGCCCTGTGGGAAGCGCTCAAAATG 1892
QY 541 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACAATTCAGGCGATGA 600
DB 1893 CCCAATCAGAGATTACCAAGAATCGGCGATGTCATGCTGAAGAACAATTCAGGCGATGA 1952
QY 601 AGCACTGCGCGCTCACCTGTGGAAGCAGCAGGAGCGCTTGGAGGCGCTTGGAGAATGAA 660
DB 1953 AGCACTGCGCGCTCACCTGTGGAAGCAGCAGGAGCGCTTGGAGGCGCTTGGAGAATGAA 2012
QY 661 TCAAGAGCTCCCGGCGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAGAGGTTGTT 720
DB 2013 TCAAGAGCTCCCGGCGCTGGAGAACTTCTGAGAGACTTTGAGCTGCAGAGGTTGTT 2072
QY 721 ACCTACCGCTCAACACTTCCCTCGCGCCACTGCAAGCGGCTCATGCACTACAAGCAGG 780
DB 2073 ACCTACCGCTCAACACTTCCCTCGCGCCACTGCAAGCGGCTCATGCACTACAAGCAGG 2132
QY 781 TCGTAGGCGGCTGTGCAAAACACACCCCGGAGCCAGCCGACTTCAGGAGTCCCGAG 840
DB 2133 TCGTAGGCGGCTGTGCAAAACACACCCCGGAGCCAGCCGACTTCAGGAGTCCCGAG 2192

QY 841 CGCTTTTGGCAGAGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 900
DB 2193 CGCTTTTGGCAGAGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGG 2252
QY 901 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTCGACAATCTTGTGG 960
DB 2253 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTCGACAATCTTGTGG 2312
QY 961 TTCGGGAAGGAGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCC 1020
DB 2313 TTCGGGAAGGAGTTTCATCGTCTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCC 2372
QY 1021 AGCAGCGATGTTCTTCTTCAACGAGCTCTCTATACAGAGCCGGGGCTGACGG 1080
DB 2373 AGCAGCGATGTTCTTCTTCAACGAGCTCTCTATACAGAGCCGGGGCTGACGG 2432
QY 1081 COTCCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 1140
DB 2433 COTCCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGA 2492
QY 1141 GCGAAGACGA 1150
DB 2493 GCGAAGACGA 2502
RESULT 4
AAC98992
ID AAC98992 standard; cDNA; 3094 BP.
XX AAC98992;
AC AAC98992;
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiac; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
XX P-PSDB; AAB54227.
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
PS Claim 1; Page 664-665; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiac and antinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for

CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ

Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;

Query Match 66.3%; Score 762; DB 21; Length 3094;
Best Local Similarity 100.0%; Pred. No. 8.5e-200;
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 GTGGTTTCAGACACAGTGTGAGCAAGAGAGCGCCATGCGCGAAGCACTGAAAGTCTCAT 448
DB 62 GTGGTTTCAGACACAGTGTGAGCAAGAGAGCGCCATGCGCGAAGCACTGAAAGTCTCAT 121
QY 449 ATTCCCGAATTTGAACCTTGCACAAATTTCTACTAATTTTCTCAAGGAATTCAGCA 508
DB 122 ATTCCCGAATTTGAACCTTGCACAAATTTCTACTAATTTTCTCAAGGAATTCAGCA 181
QY 509 ACGACTTGCCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 568
DB 182 ACGACTTGCCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 241
QY 569 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGACACTGTGGGGCTCACCTGTGGAAGCA 628
DB 242 CGATGTCATGCTGAAGAACATTCAGGGCATGAAGACACTGTGGGGCTCACCTGTGGAAGCA 301
QY 629 CAGCAGGCTTGGAGCGCTGTGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAACTT 688
DB 302 CAGCAGGCTTGGAGCGCTGTGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAACTT 361
QY 689 CTCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 748
DB 362 CTCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 421
QY 749 GCCACTGCACCGCTCATGCACTACAAAGAGGTCTCTGGAGCGGCTGTGCAACACCAACCC 808
DB 422 GCCACTGCACCGCTCATGCACTACAAAGAGGTCTCTGGAGCGGCTGTGCAACACCAACCC 481
QY 809 GCCGAGCCACGCGGCTTCAAGGACTTCCGAGCGCTTGGAGAGATCACGGAGATGTT 868
DB 482 GCCGAGCCACGCGGCTTCAAGGACTTCCGAGCGCTTGGAGAGATCACGGAGATGTT 541
QY 869 GGCACAGCTCCACGGTACGATGATCAAGATGGAATTTCCAGAACTGCACGAAGTCAA 928
DB 542 GGCACAGCTCCACGGTACGATGATCAAGATGGAATTTCCAGAACTGCACGAAGTCAA 601
QY 929 GAAAGATTGATTGGCATTTGACAACTTGTGGTTCCGGGAAGGAGTTTATCGCTGTGG 988
DB 602 GAAAGATTGATTGGCATTTGACAACTTGTGGTTCCGGGAAGGAGTTTATCGCTGTGG 661
QY 989 GAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGAGCGCATGTTCTTCTGTTCACCA 1048
DB 662 GAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGAGCGCATGTTCTTCTGTTCACCA 721
QY 1049 CGTCTGCTATACAGAGCGGGGGCTGACGGCTCCAACTCAGTTTAAAGTCCACGGCA 1108
DB 722 CGTCTGCTATACAGAGCGGGGGCTGACGGCTCCAACTCAGTTTAAAGTCCACGGCA 781
QY 1109 GTCCCGCTCTATGGCATGACGATTGAGGAGCGGAAGCA 1150
|||||

DB 782 GCTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 823

RESULT 5

AAS58814/c
ID AAS58814 standard; cDNA; 426 BP.

XX AC AAS58814;
XX

DT 13-FEB-2002 (first entry)

XX cDNA #1490 encoding portion of a human colon tumour protein.

XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.

XX Homo sapiens.

XX WO200173027-A2.

XX 04-OCT-2001.

XX 22-MAR-2001; 2001WO-US09246.

XX 24-MAR-2000; 2000US-191597P.

XX 04-MAY-2000; 2000US-202024P.

XX 05-MAY-2000; 2000US-202189P.

XX (CORI-) CORIXA CORP.

XX Meagher MJ, Xu J, King GE;

XX WPI; 2001-611627/70.

XX New colon tumour proteins and related nucleic acid, useful for treatment, prevention, diagnosis and monitoring of cancer -

XX Claim 4; Page 290; 299pp; English.

XX Th present invention relates to the isolation of novel cDNA sequences encoding for at least an immunogenic portion of human colon tumour proteins. The sequences of the invention are useful in pharmaceutical compositions and vaccines for the prevention and treatment of cancers such as colon cancer. They are also useful for the diagnosis and monitoring of such cancers. Antibodies to the colon tumour proteins and antigen presenting cells that express polynucleotides encoding colon tumour proteins can be used to inhibit the development of cancers. T-cells that react specifically with colon tumour proteins are useful for removing tumour cells from samples (e.g. blood) and for cancer treatment. The polynucleotide sequences are also useful in gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the invention that encode for portions of human colon tumour proteins.

XX Sequence 426 BP; 89 A; 109 C; 109 G; 119 T; 0 other;

Query Match 13.4%; Score 154.6; DB 23; Length 426;
Best Local Similarity 61.9%; Pred. No. 1.5e-32;
Matches 263; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 321 TACTTCATAGCTAAGGAAGTGTCTACCCAGCGAGCAATATCTGAAGATCTCGAAGTT 380
DB 426 TACTTCATAGCTAAGGAAGTGTCTACCCAGCGAGCAATATCTGAAGATCTCGAAGTT 367
QY 381 ATCACTCTCGTGTTCAGAGCACAGTGACAAAGAGAGCGCATGCCGAAGCACTGAAA 440
DB 366 ATTACCGTGTGTTCGCGAGCGAGTGTGAAGGAGAGCGCATGCCCTCGGACTCTGTG 307
QY 441 AGTTCATATTTCCGGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTCTCAAGAA 500
DB 306 ACGCTGCTCTTCTCCAAACATCGATCCATCTATAGTTCACAGAGGCTCTCGCGCAG 247
QY 501 ATTGACAAACGACTTGCCTGTGGGAAGCGCGCTCAAAATGCCCAATACAG---ATTAC 557
DB 246 GTGGAGCAGAGGCTGGCACTCTGGGAAGGGCCCTCCAAAGGCCACACAAAAGGCACTCAT 187
|||||

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-465460/50.
DR P-PSDB; AAU17099.
DR
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders -
PS Claim 1; SEQ ID No 51; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465460/50.
DR P-PSDB; AAU17523.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
XX Claim 1; SEQ ID No 475; 880pp; English.
PS
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases

(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and other blood-related disorders (sickle cell anaemia), myeloproliferative disorders, primary haematopoietic disorders, hyperproliferative disorders (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal abnormalities (down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders (e.g. psoriasis), epithelial cell proliferation, endocrine disorders (e.g. Addison's disease), reproductive system disorders, gastrointestinal disorder (inflammatory disorders), liver disorders (cirrhosis), as stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, and as a means to induce tumour proliferation in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction pathway protein coding sequences and PCR primers of the invention.

Query Match 9.1%; Score 105.2; DB 22; Length 716;
Best Local Similarity 64.0%; Pred. No. 8.5e-19;
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;
QY 909 CAGAGCTGCACGACCTCAGAAAGATTTCATTGGCATTGACAACTTGTGGTCCGGGA 968
DB 3 CAGAAGCTWACGGAGCTGCWKCGGACCTGGTGGCATAGAGAACCTCATTTGCTCTGGC 62
QY 969 AGGAGATTATCCGCTCTGGGAGCGCTCAGCAAGCTCTCGGGGAGGGCTCCAGCAGCGC 1028
DB 63 AGGAGTTATCCGCTGAGGCTGCTTCAAGCTCACCAGAAAGGCGCTGCAGCAGAGG 122
QY 1029 ATGTTCTTCTGTTCAACAGCTGCTGCTATACAGAGCGGGGGCTGACGGCTCCCAAT 1088
DB 123 ATGTTTCTGTTCTCAGATATGTTGCTGTACACAAGCAAGAGCTGTCAGGACCAGC 182
QY 1089 CAGTTTAAAGTCCAGCGGAGCTCCCGCTCTATGGCATGACGATTGAGGACCGAAGAC 1148
DB 183 CACTTCGGATCCGGGGCCCTCTCCCTCCCAAGGCATGCTGGTGAAGAAGTGATAC 242
QY 1149 GA 1150
DB 243 GA 244

RESULT 11
ABL90828
ID ABL90828 standard; cDNA; 717 BP.

AC ABL90828;

XX 24-MAY-2002 (first entry)

DE Human polynucleotide SEQ ID NO 1390.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.

OS Homo sapiens.

XX WO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001WO-US16450.

XX 19-MAY-2000; 2000US-205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX

DR WPI; 2002-122018/16.
DR P-PSDB; ABB90419.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.

XX Claim 4; SEQ ID NO 1390; 2081pp + Sequence Listing; English.

CC The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB90440-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 717 BP; 170 A; 196 C; 203 G; 144 T; 4 other;

Query Match 8.5%; Score 98; DB 24; Length 717;
Best Local Similarity 65.6%; Pred. No. 8.2e-17;
Matches 143; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 933 GATTTGATGGCATTGACAACTCTGTGGTTCGGGAAGGAGTTTCATCGTCTGGGCGAC 992
DB 28 GACCTGGTGGCATAGAGAACCTCATTTGCTCTCGCAGGAGGTTTCATCGTGGGCTGC 87

QY 993 CTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTC 1052
DB 88 CTTCAAGCTCACCAGAAAGGGCCTGCAGCAGAGGATGTTTTTCTGTTTCAGATATG 147

QY 1053 CTGCTATACAGCGCGGGGCTGACGGCTCCAACTCAATCAATCAATCAATCAATCAAT 1112
DB 148 TTGCTGTACAGCAAGCAAGAGGAGTTGCAGGAGCACCACCTCCGGATCGGGGCTCTT 207

QY 1113 CCGCTCTATGGCATGACGATTGAGGAGCGGAGACGA 1150
DB 208 CCCCTCCAAGGCATGCTGGTGAAGAAGTGATAACGA 245

RESULT 12

AAH98023

ID AAH98023 standard; DNA; 544 BP.

XX AAH98023;

XX 10-OCT-2001 (first entry)

DE Murine 7-transmembrane G-protein coupled receptor coding sequence #267.

XX Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;

XX 7-transmembrane G-protein coupled protein receptor; ds.

XX Mus sp.

XX WO200160999-A1.

XX 23-AUG-2001.

XX 14-FEB-2001; 2001WO-US04700.

XX

RESULT 14

ID ABK49898 standard; cDNA; 1296 BP.

AC ABK49898;

DT 15-JUL-2002 (first entry)

DE Human cDNA encoding Faciogenital dysplasia 1-like protein.

XX Human; ss; gene; faciogenital dysplasia; FGDL-like protein;

KW Aarskog syndrome; X-linked developmental disorder; cancer; obesity;

KW guanine nucleotide exchange factor; overweight; anorexia; cachexia;

KW wasting disorder; appetite suppression; appetite enhancement; satiety;

KW modulation of body weight; eating disorder; bulimia; hypertension;

KW type 2 diabetes; coronary artery disease; hyperlipidaemia;

KW stroke; gallbladder disease; gout; osteoarthritis; sleep apnea;

KW respiratory problem; thrombotic disease; polycystic ovarian syndrome;

KW reduced fertility; complication of pregnancy; menstrual irregularity;

KW hirsutism; stress incontinence; depression.

XX Homo sapiens.

OS

XX

FH Key Location/Qualifiers

FT CDS 1..1296

FT /*tag= a

FT /product= "FGDL-like protein"

FT /partial

FT /note= "No stop codon shown"

PN WO200224731-A2.

XX

PD 28-MAR-2002.

XX

PF 24-SEP-2001; 2001WO-EPI1009.

XX

PR 25-SEP-2000; 2000US-234983P.

XX

PA (FARB) BAYER AG.

XX

PI Kossida S;

XX

DR WPI; 2002-383181/41.

DR P-PSDB; AAU80165.

XX

New human faciogenital dysplasia (FGDL)-like proteins, regulators of which are useful for preventing and treating cancer, obesity, cachexia, depression, diabetes, hypertension and stroke -

Claim 1; Fig 1; 78pp; English.

The invention relates to a purified human faciogenital dysplasia (FGDL)-like protein and its encoding cDNA. Faciogenital dysplasia (also known as Aarskog syndrome) is an X-linked developmental disorder and is a guanine nucleotide exchange factor. Also included are a hybridisable sequence, fragment, derivative, allelic variant, or a sequence which deviates from the cDNA (or sequence 50% identical to it) due to degeneration of the genetic code, an expression vector comprising the nucleotide sequence, a host cell containing the vector, modulators of the activity of the protein, fusion proteins with the FGDL-like protein, and screening for agents which modulate/reduce the activity of the protein. The FGDL-like protein and cDNA are useful for screening for agents which decrease the activity of FGDL-like protein. The cDNA is useful for detecting a polynucleotide encoding a FGDL-like protein in a biological sample. The protein modulators and isolated agents are useful for treating a FGDL-like protein dysfunction related disease, such as cancer, obesity, overweight, anorexia, cachexia, wasting disorders, appetite suppression, appetite enhancement, increases or decreases in satiety, modulation of body weight, and/or other eating disorders such as bulimia, obesity/overweight-associated co-morbidities including hypertension, type 2 diabetes, coronary artery disease, hyperlipidaemia, stroke, gallbladder disease, gout, osteoarthritis, sleep apnea and respiratory problems, endometrial, breast, prostate, colon cancer,

CC thrombotic disease, polycystic ovarian syndrome, reduced fertility, complications of pregnancy, menstrual irregularities, hirsutism, stress incontinence and depression. The coding region of FGDL-like protein polynucleotide is useful for generating antisense oligonucleotides or ribozymes. These antisense oligonucleotides are useful for modulating FGDL-like protein gene expression. The FGDL-like protein is useful for generating antibodies against FGDL-like protein amino acid sequences and for use in various assay systems. The present sequence is the cDNA encoding the human FGDL-like protein.

SQ Sequence 1296 BP; 415 A; 273 C; 281 G; 327 T; 0 other;

Query Match 7.3%; Score 83.6; DB 24; Length 1296;
Best Local Similarity 51.3%; Pred. No. 1e-12;
Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 719 TTACTTACCCTGCAACACCTTCCTCCTCGGGCCCACTCACCGGCTCATGCATACAAGCA 778
DB 18 TAATCTGGCCCTCAAGCACTACCTGCTCAAGCGGTTCAGAGGATCCCCAGTACAGGCT 77
QY 779 GGTCCTGGAGCGGCTGTGCCAAACACCACCCGCCGAGCCGACCTTCAGGGACTGCCG 838
DB 78 GTTGCTGACAGATTATTTGAAGAATCTCATAGAAGATGCTGGAGATTACAGACACTCA 137
QY 839 AGCCGCTTTGGCAGAGATCAGGAGATGGTGGCACAGCTCCAGGTACGATGATCAAGAT 898
DB 138 AGATGCCCTTCTGCTTTATAGAGTAGCCAACCCAGCAATGACACCATGAAGCAAGG 197
QY 899 GGAGAAATTTCCAGAAGCTGCACGAACATCAAGAAAGATTGATGGCAATTGACAACTTGT 958
DB 198 AGACAACCTTTCAGAAACTTATGCAAATTCAGTACAGCTTAATGGACACCATGAATTTGT 257
QY 959 GGTTCGGGGAAGGAGTTTCATCCGCTCTGGGCGAGCTCAGCAAGCTCTCGGGGAAGGGGCT 1018
DB 258 GCAGCTGCTGGGTCTTCATCAAGAAGGAAATTCATGAGCTGCTCGAAGAGTGAT 317
QY 1019 CCAGCAGCGCATGTCTTCCTGTTTCAACAGAGCTCCTGCTATACACAGCGGGGGGTGAC 1078
DB 318 GCAACCTCGAATGTTTTCTGTTTAATGATGCCCTGCTGTATACAAACACCAAGTGCAGTC 377
QY 1079 GGCCTCCAATCAGTTTAA 1096
DB 378 TGGGATGTATAAAGTGAA 395

RESULT 15

AHH15014

ID AHH15014 standard; cDNA; 1821 BP.

XX AC AAH15014;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:12971.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 34.959 Seconds
(without alignments)
10088.332 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1352_2501
Perfect score: 1150
Sequence: 1 cgggtacaagcagcgagc.....attgaggagagcggaagacga 1150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.8	3.5	4776	US-08-852-401-1	Sequence 1, Appli
2	40	3.5	7218	US-08-232-463-14	Sequence 14, Appli
3	39.6	3.4	4722	US-08-979-608A-14	Sequence 14, Appli
4	38.4	3.3	4267	US-08-949-155-51	Sequence 51, Appli
5	38.4	3.3	4267	US-09-819-964-51	Sequence 51, Appli
6	37.2	3.2	44377	US-08-804-227C-7	Sequence 7, Appli
7	37.2	3.2	44377	US-08-804-198-1	Sequence 1, Appli
8	36.4	3.2	2824	US-09-010-928B-3	Sequence 3, Appli
9	36.2	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
10	36	3.1	2214	US-08-864-038A-1	Sequence 1, Appli
11	36	3.1	3331	US-08-864-038A-2	Sequence 2, Appli
12	36	3.1	3331	US-08-864-038A-4	Sequence 4, Appli
13	35.4	3.1	2132	US-09-552-322-1	Sequence 1, Appli
14	35.4	3.1	4137	US-09-499-984-2	Sequence 2, Appli
15	35.4	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
16	35.4	3.1	4411529	US-09-103-840A-1	Sequence 1, Appli
17	34.6	3.0	657	US-09-527-345-3	Sequence 3, Appli
18	34.6	3.0	1029	US-08-911-853-1	Sequence 1, Appli
19	34.6	3.0	1029	US-09-479-409-1	Sequence 1, Appli
20	34.6	3.0	1029	US-09-479-453-1	Sequence 1, Appli
21	34.6	3.0	4377	US-08-911-853-28	Sequence 28, Appli
22	34.6	3.0	4377	US-09-479-409-28	Sequence 28, Appli
23	34.6	3.0	4377	US-09-479-453-28	Sequence 28, Appli
24	34.6	3.0	10348	US-08-457-273B-41	Sequence 41, Appli
25	34.6	3.0	10348	US-08-556-419-13	Sequence 13, Appli
26	34.6	3.0	10348	US-09-041-886-14	Sequence 14, Appli
27	34.6	3.0	10366	US-08-246-982A-5	Sequence 5, Appli

28	34.6	3.0	10366	1	US-08-453-265-5	Sequence 5, Appli
c 29	34.4	3.0	7042	4	US-09-092-508-1	Sequence 1, Appli
c 30	34.4	3.0	7042	4	US-09-435-115-1	Sequence 1, Appli
c 31	34.4	3.0	7042	4	US-09-098-310-1	Sequence 1, Appli
c 32	34.4	3.0	7042	4	US-09-690-364-21	Sequence 21, Appli
c 33	34.4	3.0	7075	4	US-09-092-508-15	Sequence 15, Appli
c 34	34.4	3.0	7075	4	US-09-435-115-15	Sequence 15, Appli
35	34.2	3.0	1331	4	US-09-335-983-14	Sequence 14, Appli
36	34.2	3.0	4195	1	US-08-340-011-1	Sequence 1, Appli
37	34.2	3.0	4195	3	US-08-901-710-1	Sequence 1, Appli
38	34.2	3.0	4416	3	US-08-795-430-1	Sequence 1, Appli
39	34.2	3.0	4416	4	US-09-355-700-1	Sequence 1, Appli
40	34.2	3.0	4416	4	US-08-601-132-36	Sequence 36, Appli
41	34.2	3.0	4425	1	US-08-222-616-31	Sequence 31, Appli
42	34.2	3.0	4425	4	US-08-446-648-31	Sequence 31, Appli
43	34.2	3.0	4425	5	PCT-US95-04228-31	Sequence 31, Appli
44	34.2	3.0	4795	1	US-08-340-011-3	Sequence 3, Appli
45	34.2	3.0	4795	3	US-08-901-710-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-852-401-1
; Sequence 1, Application US/08852401
; Patent No. 5976836
; GENERAL INFORMATION:
; APPLICANT: Weber, J. Mark
; APPLICANT: Hessler, Paul E.
; APPLICANT: Larsen, Peter E.
; APPLICANT: Luu, B. Minh
; TITLE OF INVENTION: Methods and Compositions for Enhancing
; Erythromycin Production
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
; STREET: 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,401
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueller, Lisa L.
; REGISTRATION NUMBER: 38,978
; REFERENCE/DOCKET NUMBER: FER2159P00300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-852-401-1

Query Match 3.5%; Score 40.8; DB 2; Length 4776;
Best Local Similarity 43.9%; Pred. No. 0.083;
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 661 TCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTCGAGAGGTGTGTT 720

Db 797 TCAACCGCGGCGGCATCAAGGAGGACACCGCGCGCGTGTGGCCGTCTCAACG 856
Qy 721 ACCTACCGCTCAACACCTTCTCTCGCGGCACATCGACCGGCTCATGCACTACAGCAGG 780
Db 857 AGATCGGCACACCGCGCGGCATAGTGGCTGCGCGCACCGGGTAATCGCGCTGC 916
Qy 781 TCCTGAGCGGTGTGAAACACACCGCGGAGCGACCTTCAGGGACTGCCGAG 840
Db 917 TGGTCCGAGAGTGTCAACCGGTGTCGCGGCTTCGCGAGGCCCTGGAGGCGCGC 976
Qy 841 CGCTTTGCGAGATACCGGAGATGGTGGACAGCTCCACGGTACGATGATCAAGATGG 900
Db 977 CGTCCGCGCGGTACGCTCGCTGCTGTCAACACCGCGCTCGGATGATGCGGAGG 1036
Qy 901 AGAATTTCCAGAGCTGCAGCACTCAAGAAGATTTGATGGCAATTCACAAATCTTGG 960
Db 1037 ACTAGTCGGATGCTATCGCCCGCGCGTGGAGGATGCTTCGTGTCGCGCGAGA 1096
Qy 961 TTCGCGGAGGAGTTCATCGTCTGGGCGAGCTCAGCAAGCTCTCGGGGAGGGGCTCC 1020
Db 1097 TCGCCAAACCGAGGCGGAGCGGATCAGCGCGAGCTACTACAGAAGCTGTGGCGC 1156
Qy 1021 AGCAGCGCATGTTCTCTCTGTTCACAGCAGTCTCGC 1056
Db 1157 ACGGCGTGGCATGCTCTCGTCAACGCGCGCGC 1192

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; IMMEDIATE SOURCE:
; CLONE: ptzgpT-Fls
US-08-232-463-14
Query Match 3.5%; Score 40; DB 1; Length 7218;
Best Local Similarity 5.4%; Pred. NO. 0.18;
Matches 22; Conservative 206; Mismatches 176; Indels 0; Gaps 0;
Qy 280 AGGATGAGGCGCGGAGGAGATTCACCACTGATAAGCGTACTTCATAGTAAGGAG 339
Db 1446 AGAATTTGGTACRR 1387
Qy 340 TGTCTACCACGAGGAGACATATCTGAAGATCTCGAAGTTATCACTTCGTGGTTTCA 399
Db 1386 RRR 1327
Qy 400 GCACAGTGAGCAAGAGAGCGCCATCGCGAAGCACTGAAAGTCTCATATTCGCCAAT 459
Db 1326 RRR 1267
Qy 460 TTGAACCTTTCACAAATTTTCATACTAATTTCTCAAGAAATTCAGCAACGACTGCC 519
Db 1266 RRR 1207
Qy 520 TGTGGAAGCGCGTCAAAATGCCAAATCAGAGATTACCAAGATCGCGATGTCATGC 579
Db 1206 RRR 1147
Qy 580 TGAAGAACATTCAGGCGCATGAAGCACCTCGCGGCTCACTGTGGAAGCACGCGCCT 639
Db 1146 RRR 1087
Qy 640 TGGAGCGCTCGAATGGAATCAAGACTCCCGCGCGCTGGAG 683
Db 1086 RRR 1043

RESULT 3
US-08-979-608A-14
; Sequence 14, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Law, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965

```
;
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 61...1731
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-979-608A-14

Query Match          3.4%; Score 39.6; DB 4; Length 4722;
Best Local Similarity 47.5%; Pred. No. 0.18;
Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;

QY 599 GAAGCACCTGGCGGCTCACCTGTGAAGCACAGCAGAGCGCCTTGGAGCCCTGGAGAATGG 658
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 GGAGAAGCTGGCGCTCTGTGCAAGAGTATGCGGAAGTCTCTCGAGGAGCACCAGGAACTC 677

QY 659 AATCAGAGCTCCGGGGCTGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAAAGTGTG 718
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 GCAGAAGCAGATGAAGTGTCTGCAGAAGAGCAGAGCCAGCTGTGTGAGGAGAAGG--- 733

QY 719 TTACCTACCGCTCAACACCTTCCTCTCGGGCCACCTGCACCGGCTCATGCACCTACAAGCA 778
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 --ACCACCTCGTGGAGCAGACAGCAAGGCCATCTCTGGCCCGCAGCAAGCTCGAGAGCCT 791

QY 779 GGTCTGGAGCGGCTGTGCAAAACACACCGCGAGCCGCGACTTCAGGGAGCTGCCG 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 GTCCGGGAGCTCGAGCGGCACAAACCGCTCGCTCAGGAGAAGTGTGCAGCGAGCCCG 851

QY 839 AGCCGCTTTGGAGAGATCAGGAGATGTTGGCAGCTCCAGCTCCAGTATCAAGAT 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 AGAGAGGAGGAGAGCGGAAGAGGTGAGTCAACACTTCCAGATGACGCTCAACGACAT 911

QY 899 GGAGAAATTTCCAGAAGCTGCACGA 922
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 TCAGTGCAGATGGAGCAGCACAA 935

RESULT 4
US-09-819-964-51
; Sequence 51, Application US/08949155
; Patent No. 6271436
; GENERAL INFORMATION:
; APPLICANT: Piedrahita, Jorge A
; APPLICANT: Bazer, Fuller W
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Generation of Transgenic Animal Species
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/949,155
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/027,338
;
```

```
;
; FILING DATE: 11-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,094
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Habler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: TAMK:177
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-949-155-51

Query Match          3.3%; Score 38.4; DB 4; Length 4267;
Best Local Similarity 47.2%; Pred. No. 0.39;
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 24 GCCGCCCTCGCGCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3020 GCCCTCTCTCGCGCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3079

QY 84 AAACCTCAGCCCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3080 GAGGAGCTGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3139

QY 144 CTGCTCTGAACTCGCAGGGGGAGTGGCCCTCGCCAAAGTGACCTTGCTTCCCAACCTG 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3140 GAGCTCGAGGCGCGAGGCGCCGCTGGGCGCGAGCATGGAGGAGCTGCGCAACCGCTTG 3199

QY 204 AGCCCGCAGCACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTGCTGATGACCAAGGCTG 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3200 GTGCTCTACCGCAGGAGGTGCAACAACATGTTGGCGCAGCACCAAGGAGGAGCTGCGGAG 3259

QY 264 CCCCAGGAC 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3260 CGCCTGGC 3267

RESULT 5
US-09-819-964-51
; Sequence 51, Application US/09819964
; Patent No. 6369294
; GENERAL INFORMATION:
; APPLICANT: Piedrahita, Jorge A
; APPLICANT: Bazer, Fuller W
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Generation of Transgenic Animal Species
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/819,964
; FILING DATE: 28-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,155
; FILING DATE: <Unknown>
;
```



```

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          3.1%; Score 36.2; DB 4; Length 4403765;
Best Local Similarity 46.6%; Pred. No. 41;
Matches 116; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY   27 GCCTCGGCGCCCGAGGAGGAGGAGGTCTTAAGATAGGACCCAGCAGATATA 86
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   3896461 GCCTGTGTGGCTGCCCGGTCGTTGCGTACGCGGTATGCATCGGGTCCGACCAGGTGCC 3896402

QY   87 CCTCACGCCCCCGACGCAACACAGAGCTCCTTGACTGGCAGTCCCTCACCTTTCCGAGGTG 146
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   3896401 AACACGGCGCGCGGCAACGACGTCCGACGCGCGCGCCCACATTCCCCGGCATCCAG 3896342

QY   147 TCTGTGAAGTCCGAGGGGGAGTGGCCCTTGCAACGTGACCTTGTCTCCCAACCTGAGC 206
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   3896341 GTTGGCCCGTGTGGCGCGGACGAGGCAACGGCCGCTGCGCGCGCTGGCCAGCTTGCC 3896282

QY   207 CCAGACACCAAGCAGGCGCTCTCCCTTGATCAGCCGCGCTGCTGAATGACAGCGCTGCCCC 266
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   3896281 CCGGATTCAAACACGGCAGGCGGTGGTAATGATCGCCGGAGGACCACTACGACCCG 3896222

QY   267 CGGACGGAC 275
      | |||||
Db   3896221 TGCGCGGAC 3896213

RESULT 10
US-08-864-038A-1/c
; Sequence 1, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: KUNIO NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 37.9989 Seconds
(without alignments)
11655.134 Million cell updates/sec

Title: US-09-555-342b-1_copy_1352_2501

Perfect score: 1150

Sequence: 1 cgggtacaagcagcgacgac.....attgaggagcggaagacga 1150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	762	66.3	3094	10	US-09-925-297-220
2	297	25.8	379	10	US-09-960-352-154
3	292	25.4	400	10	US-09-960-352-156
4	248.6	21.6	408	10	US-09-783-590-3620
5	237.4	20.6	500	10	US-09-783-590-3621
6	190.8	16.6	335	10	US-09-783-590-3600
7	168.8	14.7	356	10	US-09-783-590-3575
8	154.6	13.4	426	10	US-09-815-343-1490
9	133.2	11.6	288	10	US-09-783-590-3542
10	106	9.2	1718	9	US-09-764-868-51
11	105.2	9.1	1716	9	US-09-764-868-475
12	64.4	5.6	2686	9	US-09-764-868-48
13	64.4	5.6	3380	10	US-09-799-799-1
14	64	5.6	1091	9	US-09-764-868-316
15	43.4	3.8	3786	10	US-09-815-242-7865
16	42.4	3.7	1730	9	US-09-860-670-78
17	42.4	3.7	1730	9	US-09-764-868-52
18	42.4	3.7	1730	9	US-09-764-868-476
19	40.8	3.5	2172	10	US-09-815-242-4038

c	20	40.4	3.5	598	10	US-09-770-149-991	Sequence 991, App
	21	40.4	3.5	768	9	US-09-938-842A-812	Sequence 812, App
	22	39.6	3.4	4722	10	US-09-962-055-14	Sequence 14, Appl
	23	39.6	3.4	4722	12	US-10-023-529-14	Sequence 14, Appl
	24	39.6	3.4	4722	12	US-10-023-523-14	Sequence 14, Appl
c	25	39.4	3.4	420	10	US-09-960-352-9773	Sequence 9773, Ap
	26	38.4	3.3	1685	9	US-09-954-531-133	Sequence 133, App
	27	38.4	3.3	1685	9	US-09-954-531-354	Sequence 354, App
	28	38.4	3.3	1685	10	US-09-962-436-261	Sequence 261, App
	29	38.4	3.3	2296	10	US-09-822-849A-259	Sequence 259, App
c	30	37.8	3.3	420	10	US-09-960-352-9532	Sequence 9532, Ap
	31	37.8	3.3	422	10	US-09-738-973-337	Sequence 337, App
c	32	37.4	3.3	552	9	US-09-764-868-406	Sequence 406, App
	33	37.4	3.3	16747	10	US-09-764-877-3354	Sequence 3354, Ap
	34	36.6	3.2	273	10	US-09-864-761-17066	Sequence 17066, A
	35	36.6	3.2	403	10	US-09-960-352-11685	Sequence 11685, A
	36	36.6	3.2	1369	10	US-09-443-704-17	Sequence 17, Appl
	37	36.2	3.1	866	10	US-09-764-870-262	Sequence 262, App
	38	36.2	3.1	3121	9	US-10-033-245-6	Sequence 6, Appli
	39	36.2	3.1	3121	12	US-10-033-223-6	Sequence 6, Appli
	40	36.2	3.1	3121	12	US-10-033-246-6	Sequence 6, Appli
	41	36.2	3.1	3121	12	US-10-033-301-6	Sequence 6, Appli
	42	36.2	3.1	3121	12	US-10-033-326-6	Sequence 6, Appli
c	43	36	3.1	1100	10	US-09-263-959-271	Sequence 271, App
	44	36	3.1	4966	10	US-09-764-869-1885	Sequence 1885, Ap
	45	36	3.1	7238	10	US-09-764-869-1886	Sequence 1886, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-220
; Sequence 220 Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 220
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-220

Query Match	66.3%;	Score 762;	DB 10;	Length 3094;
Best Local Similarity	100.0%;	Pred. No. 2.4e-209;		
Matches 762;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	389	GTGGTTTCAGACACAGTCAGCAAGAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT	448	
Db	62	GTGGTTTCAGACACAGTCAGCAAGAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT	121	
QY	449	ATTCCCGAATTTTGACCTTTGCACAAATTTTCTACTAATTTTCTCAAGAAATTCAGCA	508	
Db	122	ATTCCCGAATTTTGACCTTTGCACAAATTTTCTACTAATTTTCTCAAGAAATTCAGCA	181	
QY	509	ACGACTTGCCCTGTGGGAAGCGCGCTCAAAATGCCCAATACAGAGATTACCAAGAAATCGG	568	

Db	309	GAAGAACATTACCGGATGAAGCAACTGAGCTGCTCATTGTGGAAGCACAGCTAGGCC	671	
Qy	640	TGGAGGCCCTGGAGATGGAATCAAGAGCTCC	400	
Db	369	TGGAGGCCCTGGACATCGGCATCTGGGCC	400	
RESULT 4				
US-09-783-590-3620				
; Sequence 3620, Application US/09783590				
; Patent No. US20020110850A1				
; GENERAL INFORMATION:				
; APPLICANT: Dillon, Patrick J.				
; APPLICANT: Haseltine, William A.				
; APPLICANT: Li, Haodong				
; APPLICANT: Rosen, Craig A.				
; APPLICANT: Ruben, Steven M.				
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16				
; FILE REFERENCE: PO-16.2C1				
; CURRENT APPLICATION NUMBER: US/09/783,590				
; CURRENT FILING DATE: 2000-02-15				
; PRIOR APPLICATION NUMBER: 08/420,856				
; PRIOR FILING DATE: 1995-04-12				
; PRIOR APPLICATION NUMBER: 08/346,731				
; PRIOR FILING DATE: 1994-11-21				
; NUMBER OF SEQ ID NOS: 12485				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 3620				
; LENGTH: 408				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
; FEATURE:				
; NAME/KEY: misc feature				
; LOCATION: (73)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (320)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (340)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (352)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (356)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (399)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (402)				
; OTHER INFORMATION: n equals a,t,g, or c				
; NAME/KEY: misc feature				
; LOCATION: (406)				
; OTHER INFORMATION: n equals a,t,g, or c				
US-09-783-590-3620				
Query Match 21.6%; Score 248.6; DB 10; Length 408;				
Best Local Similarity 85.0%; Pred. No. 5.1e-62;				
Matches 347; Conservative 0; Mismatches 46; Indels 15; Gaps				
Qy	420	GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCAAATTTGAACCTTTGCACAAATTT	479	
Db	1	GGCAGAGCGGAGCACTGAAAAGTCTCATATTCGCCAAATTTGAACCTTTGCACAAATTT	60	
Qy	480	CATACTAAATTTCTCAAGGAAATTGAGCAACGACTTGCCTCTGTGGGAAGCGCGCTCAAAAT	539	
Db	61	CATACTAAATTTTCAAGGAAATTGAGCAACGACTTGCCTCTGTGGGAAGCGCGCTCAAAAT	120	
Qy	540	GCCCAATCAGAGATTACCAAGAATTCGGGATGTGTCATGCTGAAGAACATTCAGGCGATG	599	

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (436)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (450)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (459)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (461)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (475)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (499)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3621

Query Match 20.6%; Score 237.4; DB 10; Length 500;
Best Local Similarity 83.3%; Pred. No. 9.4e-59;
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13:
QY 420 GCATGCGCGGAGCACTGAAAAGTCTCATATTCGCCGAAATTTTGAACCTTTGCACAAATTT 479
Db 1 GCGAGAGCGGAAGCACTGAAAAGTCTCATATTCGCCGAAATTTTGAACCTTTGCACAAATTT 60
QY 480 CATACTAATTTTCTCAAGGAAATGAGCAACGACTTGGCCCTGT-GGGAAGGCGGCTCAAA 538
Db 61 CATACTAATTTTNTCAAGGAAATGAGCAACGACTTGGCCCTGTGGGAAGGCGGCTCAAA 120
QY 539 TCGCCAAATCAGAGATTACCAAGATCGCGGATGTCATGCTGAA-GAACATTCAGGCA 597
Db 121 TCGCCAAATCAGAGATTACCAAGANTCGCGGATGTCATGCTGAAAGGAACATTCAGGCA 180
QY 598 TGAAGCACCTGCGGG-CTCACCTGT-GGAAGCACAGCG-AGGCCCTTGGAGGCCCTGGAGA 654
Db 181 TGAAGCACCTGCGGGTCTCACCTGTGGGAAGCACAGGAGGCCCTTGGAGGCCCTGGAGN 240
QY 655 A--TGAATCAGAGCTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 709
Db 241 AATGGNATTCAGAGCTCCCGCGGCTGGAGGANNCTTCTGCAGAGACTTTGGAGCTGCA 300
QY 710 GAAGGTGTG-TTACCTACCGCTCACACCTT--CCTCCTGCGGCCACTCCACCGGCTCAT 766
Db 301 GAAGGTGTGTTTACCTACCGCTCACACCTTCCCTNCTGCGGGCCACTGCACCGGTTTAT 360
QY 767 GCACCTACAAG--CAGTCTCTGGAGGGCTGTGCAAAACACACCCGCGAGCCACGCCGA- 823
Db 361 GCACCTACAGGCAAGGNTCTNGAGGGGNTTTCRAAAGAACACCCGATCGNGGCCAAG 420
QY 824 -----CTTCAGGAGTCCGCGCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCT 877
Db 421 GCCGATTTTCAGGANTCCGAGCGG-TTTNGCAGNNTTNGAGGATTTCTTGGGANANTT 479
QY 878 CCACGGTACGATGATCAA 895
Db 480 CCAGGGTNGTGTGTTCAA 497

RESULT 6
US-09-783-590-3600
; Sequence 3600, Application US/097833590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3600
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (29)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (36)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (157)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (184)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (254)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)


```

US-09-815-343-1490/c
; Sequence 1490, Application US/09815343
; Patent No. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ. ID NOS: 1556
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ. ID NO 1490
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-1490

```

Query Match 13.4%; Score 154.6; DB 10; Length 426;
Best Local Similarity 61.9%; Pred. No. 5.9e-35;
Matches 263; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

	Qy	321	TACTTCATAGCTAAGGAAGTGCTTACCA	CAGCGCAACATATCTGAAGGATCTCGAAGTT	380
	Db	426	TACTTCATAGTCAAAGAGATTCTCGCT	CAGAAAGCAACATACTCAAGGATTTAGAAGTT	367
	Qy	381	ATCAGCTTCGTGGTTTCAGAGCACAGT	GAGCAAAAGAGGCCCATGCCGGAAGCACTGA	440
	Db	366	ATTACCGTGTGGTTCCGACGCAGTGGT	GAGGAGAGCGCCATGCGTCCGACTCTGATG	307
	Qy	441	AGTCTCATATTTCCGAATTTTGAACCTT	TGCAAAATTTTCATACTAATTTTCTCAAGGA	500
	Db	306	ACGTGTCTTCTCCAACATCGATCCATC	TATGAGTTCCACAGAGGCTTCTGTGCGCGAG	247
	Qy	501	ATTGAGCAAGACTTGCCTGTGGGAAGCC	CGCTCAATGCCCAATTCAGAG---ATTAC	557
	Db	246	GTGGAGCAGAGGTGGCACTCTGGGAAG	GGCCCTCCAAAAGCCCAACAAGAGCAGTCA	187
	Qy	558	CAAGAATCGCGGATGTCATGCTGGAAGA	ACATTCAGGGCATGAAGCACCTGGCGGCTAC	617
	Db	186	CAAGAAATCGGCACATCCTTGCTCAGSA	AACATGCGCCAGTTAAAGAGGTTACCAGCTAC	127
	Qy	618	CTGTGNAGCACAGCAGGCGCTTGGAGG	CCCTTGGAGAAATGGAAATCAAGAGCTCCC	677
	Db	126	TTCCAAGACATGACGAGATCTCTAACAG	AACTGTGAAAGAGGCTACCAAAAGCTGTAGA	67
	Qy	678	CTGGAGAACTTCTCGAGAGACTTTGAGCT	GCAAGGTTGTTTACCTACCGCTCAACACC	737
	Db	66	TTGGAGGCGAGTGTACRAGGAGTTTGAG	CTGCGAGAGGTCGCTACTTGCCCTCTCAACAG	7
	Qy	738	TTCTT 742		
	Db	6	TTCTT 2		

RESULT 9

```

US-09-783-590-3542
; Sequence 3542, Application US/09783590
; Patent No. US20030110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Polymers
; FILE REFERENCE: PO-16-2C1
; CURRENT APPLICATION NUMBER: US/09/783
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12

```

APPLICANT: RUBEN, STEVEN M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2

```

; FILE REFERENCE: PO-16.2CI
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12

```

```

: PRIOR APPLICATION NUMBER: 08/346,731
: PRIOR FILING DATE: 1994-11-21
: NUMBER OF SEQ ID NOS: 12485
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3542
: LENGTH: 288
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (3)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (42)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (72)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (75)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (93)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (129)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (142)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (155)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (156)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (166)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (168)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (182)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (253)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (263)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (264)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (267)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (270)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (272)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (282)
: OTHER INFORMATION: n equals a,t,g, or c
: US-09-783-590-3542

```

Query Match 11.6%; Score 133.2; DB 10; Length 288;
Best Local Similarity 85.8%; Pred. No. 7e-29;
Matches 188; Conservative 0; Mismatches 27; Indels 4;

QY 427 CGGAAGCACTGAAAAGCTCATATTTCCCGAAATTTGAACCTTTGCACAAATTTTCATACTA 486


```

Db      8 CGGAAGCAGCTGAAAGTGTCAAAATCCCAAATTTNAAACCTTTGCACAAATTAAACTA 67
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy     487 ATTTCCTCAAGGAAATTGAGCAACGACTTGCCTCTGTGGAGGCGGCTCAAATGCCCAA 546
        ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     68 ATTTNTNRAAGGAAATTAAGCAACGNCITGCCCCTGTGGNAAGCGCCTCAAATGCCCAA 127
        ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy     547 TCAGAGATTACCAGAAGATCGGCGATGT-CATGCTGAAG-AACAATCAGGCATGAAGCA 604
        ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    128 TNA-AGATTACCAANAATCGCGATGTNNATGCTGAAGNANCAATTCAGGCGCATNGSCA 186
        ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy     605 CCTGGCGGC-TCACCTGTGGAACACAGCGAGGCCTTGG 642
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    187 CQTGGCGGCTTCACCTGTGGGAAGCACGCGAGGCGCTTG 225
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-764-868-51
; Sequence 51, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-51

```

```

Query Match          9.28; Score 106; DB 9; Length 1718;
Best Local Similarity 64.9%; Pred. No. 1.2e-20;
Matches 157; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY  909  CAGAAGCTGCACGAACTCAAGAAAGATTGATGGCATTTGCAATCTTGTGTTCCGGGA  968
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db   13  CAGAAGCTAACGGAGCTGCACGGGACCTGTGGCATAGACAACCTCATTTGCTCTCGGC  72

QY  969  AGGAGGTTCAATCCCTCTGGGCGAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGC  1028
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db   73  AGGAGGTTCAATCCCTGTAGGGCTGCTTTCACAAGCTCACAAAGAGGGGCTGCAGCAGAGG  132

QY  1029  ATGTTCTTCTCTGTTCAACGAGCTCTCTGTATACAGAGCCGGGGGCTCACGGCCTCCCAAT  1088
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db   133  ATGTTTTTTCGTCTCTCAGATATGTTGCTGTACACAAGCAAGGAGTTGCAGGACCAGC  192

QY  1089  CAGTTTAAAGTCCACGGGCAGCTCCCGCTCTATGGCATGAGATTTAGAGGAGGGAAGAC  1148
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db   193  CACTTCCGGATCCGGGGCCTCTTCCCTCCCAAGGCATGCTGTGTGAAGAAGTGATAAC  252

QY  1149  GA 1150
      ||  ||
Db   253  GA 254

RESULT 11
US-09-764-868-475
; Sequence 475, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ-ID NO 475
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (715)
; OTHER INFORMATION: n equals a,t,g, or c
us-09-764-868-475

Query Match 9.1%; Score 105.2; DB 9; Length 716;
Best Local Similarity 64.0%; Pred. No. 1.3e-20;
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;

Qy 909 CAGAAAGTCGACGAACTCACAAGAAAGATTGATGGCATTTGACAATCTTGTGGTTCCGGGA 968
Db 3 CAGAAAGCTWACGGAGCTCWCKGGGAACTGTGGGCATAGAGAACCTCATTTGCTCTGGC 62

Qy 969 AGGAGATTCAATCCGTCTGGGAGGCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGGCG 1028
Db 63 AGGAGATTCAATCCGTGAGGGCTGCTTACAAAGCTCACAAGAGGGGCTCGACGAGAG 122

Qy 1029 ATGTTCTTCCTGTTCAAGCAGGTCTGTGTATACAGAGCCGGGGCTGACGGCTTCCAAT 1088
Db 123 ATGTTTTTCTGTTCTCAGATATGTTGCTGTACACAAGCAAGGAGTTGTCAGGACCACG 182

Qy 1089 CAGTTTAAAGTCCACGGGACGCTCCGCTCTATGCGATCGACGATTGAGGAGAGCGGAAGAC 1148
Db 183 CACTTCGGATCCGGGGCTCTCTTCCCTCCAGGCATGCTGGTGGAAAGATGATAAC 242

Qy 1149 GA 1150
Db 243 GA 244

```

```

RESULT 12
US-09-764-868-48
; Sequence 48, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2505)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2569)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2644)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2645)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-48

Query Match      5.6%   Score 64.4; DB 9: Length 2686;
Best Local Similarity 47.8%; Pseq No. 1.5e+08;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

Qy 556 ACCAAAGAATCGGCGATGTTCATGCTGAAGACATTTCAGGGCATGAAGCACCTGGCGGCTC 615
|||||

```

```
Db 1505 ACCACGGCTCGGGACATCTCGAGAAGCTGGCCCACTTCCTGAAAGATGTACGGCGAGT 1564
QY 616 ACCTGTGGAGACACACGAGGCTTCGAGCCCTCGAGATCGAATCAAGAGCTCCCGGC 675
Db 1565 ATGCTCAAGAACTTTGACCGAGCCGTAGGCTGTGAGCAGTGGACCCAGCGCTCCCAAC 1624
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAAAGGTGT---GTTACCTACCGCTCA 732
Db 1625 TGTTTAAAGACGTCRTCCACAGATCCAGAAACAGGAGGTATCGCGGAACCTGACGCTCC 1684
QY 733 ACACCTTCCTTCGCGGCCTCACTGCACCGCTCATGCACTACAAAGCAGGTCTCGGAGCGGC 792
Db 1685 AGCACCACATGCTGGAGCCCGTGCAGAGGGTCCCGCGTACGAGCTGCTCTCAAGGACT 1744
QY 793 TGTGCAAAACACCCCGCAGCCAGCCGACTTCAGGGACTGCCAGCGCTTTGGCAG 852
Db 1745 ATCTGAAGAGGCTTCCCGCAGAGCGCCCAAGACCGGAAGGATCGGAGAGGTCTTTGGAGC 1804
QY 853 AGATCACGGAGATGTTGGCAGCTCCACGGTACGATGATCAAGATGGAGAATTTCCAGA 912
Db 1805 TCATCTCCACAGCGCCCAACCACTCCAATGCTGCCATTCGGAAGTGGAGAAATGCACA 1864
QY 913 AGCTGCACGAACTCAAGAAAGATTTGATTTGGCAATGACAAATCTTTGTTTCGGGAAGGG 972
Db 1865 AGCTCTTGGAGGTGTACGAGCAGCTGGTGGGAAGAAGACATTTGTCAACCGCGCAATG 1924
QY 973 AGTTCACTCGTCTGGCAGCCCTCAGCAAGCTCTCGGGGAG-----GGGCTCCAGCAGC 1026
Db 1925 AACTGATCAAGGAGGGCCAAATCCAGAACTGTACGCCAAGAACGCAACCCCGCAGGACC 1984
QY 1027 GCATGTTCTTCTTCAACGAGCTCTGCTATPACAGCGCGGGGCTGACGGCC---T 1083
Db 1985 GCCACCTCTTCTTCAACAGCATGATCTTTACTGTGTCACAGCTGGCGCTCATGG 2044
QY 1084 CCAATCAGTTTAAAGTCCAGCGGCAGCTCCCGCTCTATGCAATGATGAGGAGAGCG 1143
Db 2045 GCCAGAAGTTTCAGCTCCGCGGAGAAGATGGACATCTCAGGCTCCAGGTCAGGATATCG 2104

RESULT 13
US-09-799-799-1
; Sequence 1, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS.
; FILE REFERENCE: CL001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3380
; TYPE: DNA
; ORGANISM: Human
US-09-799-799-1

Query Match 5.6%; Score 64.4; DB 10; Length 3380;
Best Local Similarity 47.8%; Pred. No. 1.7e-08;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 556 ACCAAAGATCGCGGTATGTCATCTCTGAAGNACATTCAGGCGATGAAGCACCCTGGCGGCTC 615
Db 1225 ACCACGGCTCGGGAGACATCTCTCAAGAAGCTGGCCCCATCTCTGAAAGATGTACGGCGAGT 1284
QY 616 ACCTGTGGAAGCACAGCGAGCCCTTGGAGCCCTGGAGAAATCAAGAGCTCCCGGC 675
Db 1285 ATGTCAGAAGACTTTGACCGAGCCGTAGGCTGTGAGCAGCTGGACCCAGCGCTCCCAAC 1344
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAAAGGTGT---GTTACCTACCGCTCA 732
```

```
Db 1345 TGTTTAAAGACGCTCGTCCACAGCATCCAGAAGCAGAGGATATCGGGAACTGACGCTGC 1404
QY 733 ACACCTTCTCTCTCGCGCCACTGCACCGGCTCATGCACCTACAGCAGGTCCTTGGAGCGC 792
Db 1405 AGCACCACATGCTGGAGCCCGTGCAGAGGGTCCCCCGGTACGAGCTGTGCTCAAGGACT 1464
QY 793 TGTGCAAAACACACCCCGCAGCCAGCTTCAGGGACTGCCAGCGCTTTGGCAG 852
Db 1465 ATCTGAAGAGGCTCCCGCAGGAGCGCCCAAGACCGGAAGGATCGGAGAGGTCTTTGGAGC 1524
QY 853 AGATCACGGAGATGTTGGCAGCAGCTCCACGGTACGATGATCAAGATGGAGAATTTCCAGA 912
Db 1525 TCATCTCCACAGCGCCCAACCACTCCAATGCTGCCATTCGGAAGTGGAGAAATGCACA 1584
QY 913 AGCTGCACCAACTCAAGAAAGATTTGATTTGGCAATTCACAATCTTGTGTTTCGGGAAGGG 972
Db 1585 AGCTCTTGGAGGTGTACGAGCAGCTGGGTGGGAAGAAGACATTTGTCAACCGGCGCAATG 1644
QY 973 AGTTCACTCGTCTGGCAGCCCTCAGCAAGCTCTCGGGGAG-----GGGCTCCAGCAGC 1026
Db 1645 AACTGATCAAGGAGGGCCAAATCCAGAACTGTACGCCAAGAACGCAACCCCGCAGGACC 1704
QY 1027 GCATGTTCTTCTTCAACGAGCTCTGCTATPACAGCGCGGGGCTGACGGCC---T 1083
Db 1705 GCCACCTCTTCTTCAACAGCATGATCTTTACTGTGTGCCAAGCTGCGCTCATGG 1764
QY 1084 CCAATCAGTTTAAAGTCCACGGCAGCTCCCGCTCTATGCAATGAGGATTTAGGAGAGCG 1143
Db 1765 GCCAGAAGTTTCAGCTCCGCGGAGAAGATGGACATCTCAGGCTCCAGGTCAGGATATCG 1824

RESULT 14
US-09-764-868-316
; Sequence 316, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 316
; LENGTH: 1091
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1059)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-316

Query Match 5.6%; Score 64; DB 9; Length 1091;
Best Local Similarity 47.7%; Pred. No. 1.2e-08;
Matches 286; Conservative 1; Mismatches 301; Indels 12; Gaps 3;

QY 556 ACCAAAGATCGCGGTATGTCATCTCTGAAGAACAATTCAGGCGATGAAGCACCCTGGCGGCTC 615
Db 188 ACCACGGCTCGGGACATCTCTCGAAGAGCTGGCCCCATCTCTGAAAGATGTACGGCGART 247
QY 616 ACCTGTGGAAGCACAGCGAGCCCTTGGAGCCCTTGAGAGCTTGAGAAATCAAGAGCTCCCGGC 675
Db 248 ATGTCAGAAGACTTTGACCGAGCCGTAGGCTGTGAGCAGCTGGACCCAGCGCTCCCAAC 307
QY 676 GGCTGGAGAACTTCTCGAGAGACTTTGAGCTGCAGAAAGGTGT---GTTACCTACCGCTCA 732
Db 308 TGTTTAAAGACGTCRTCCACAGCATCCAGAAAGCAGGAGGTATCGGGAACCTGACGCTGC 367
QY 733 ACACCTTCTCTCTCGCGGCCTGACCGGCTCATGCACTACAAAGCAGGTCCTTGGAGCGGC 792
Db 368 AGCACCACATGCTGGAGCCCGTGCAGAGGGTCCCGCGGTACGAGCTGTGCTCAAGGACT 427
```

QY 793 TGTGCAAAACACCCCGCGAGCCAGCCGACTTCAGGGACTCCCGAGCGCTTTGGCAG 852
Db 428 ATCTGAAGAGGCTCCCGCAGGACGCCACAGCCGGAAGGATCGGAGAGGTCTCTTGGAGC 487
QY 853 AGATCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTCAGA 912
Db 488 TCATCTCCACAGCGCGCAACCACTCCCAATGCTGCCATTCGGAAGATGGAGAAATCCACA 547
QY 913 AGCTGACGCAACTCAAGAAAGATTTGATTGGCATTGACAAATCTTGGTTCCCGGGAAGG 972
Db 548 AGCTTTGGAGGTGTACGACAGCTGGTGGGAAGAAGACATTTGCAACCCGCGCAATG 607
QY 973 AGTTATCCGCTCTGGGAGCGCTCAGCAAGCTCTCGGGGAAG-----GGGCTCCAGCAGC 1026
Db 608 AACTGATCAAGGAGGGCCAAATCCAGAAACTGTCAAGCAAGACGGCACCCCCAGGACC 667
QY 1027 GCATGTTCTTCCTGTCAACAGAGCTCCTGCTATACAGAGCCGGGGGCTGACGGCC---T 1083
Db 668 GCCACCTCTTCCTGTTCACAGCATGATCCTTTACTGTGTGCCCAAGCTCGGCTCATGG 727
QY 1084 CCAATCAGTTTAAAGTCCACGGGCGAGTCCCGCTCTATGGCATGACGATTCAGGCGATCG 1143
Db 728 GCCAAGATTACGGTCCGGGAGAAGTGCACATCTCAGGCTCCAGGTCAGGATATCG 787

RESULT 15
US-09-815-242-7865
; Sequence 7865, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7865
; LENGTH: 3786
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3786)
US-09-815-242-7865

Query Match 3.8%; Score 43.4; DB 10; Length 3786;
Best Local Similarity 47.9%; Pred. No. 0.02;
Matches 125; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 623 GAAGCACAGCGAGCCCTTGGAGGCCCTTGGAGATCGAATGGAATCAAGAGCTCCCGCGCGGTGGA 682
Db 2541 GGAGCAGGACCTTGTACGGTGCCTGCTGATGACACCCCGAGCTGCGCGAGCC 2600
QY 683 GAACCTTCTGCAGAGACTTTGAGCTGCAGAGGTGTGTACTACCGCTCAACACCTTCCT 742
Db 2601 GTTCGGCGCGCAGCGGCATGACTGGAAGAAAGGCGAGTGCAGCCGCTGCCGGGACGCAA 2660
QY 743 CCTGCGGCCACTGCACCGGCTCATGACACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACA 802
Db 2661 CCTGCCGAGCTGCACCTGTCTGAGCGCGACTACCCGAAAGTCTACCGCAAGTTCACCTC 2720
QY 803 CCACCCCGGAGAGCCCGGACTTCAGGGACTCCCGAGCGCTTTGGCAGAGATCAACGGA 862
Db 2721 GCTCGGTCCGCTGCTGGACAAGCTTGGGCAACGGCGGCAAGGGCATCGGCTGGAACACCGA 2780
QY 863 GATGGTGGCACAGCTCCACGG 883
Db 2781 GAAGGAAGTGAAGCTGGTCGG 2801

Search completed: December 3, 2002, 22:26:32
Job time : 63.9989 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 16:31:10 ; Search time 1238.16 seconds
(without alignments)
15042.359 Million cell updates/sec

Title: US-09-555-342B-1_COPY_1352_2501

Perfect score: 1150
Sequence: 1 cgggtacaacgagcgac.....attgaggagcggaagacga 1150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estim:*
4: em_estimu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfum:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_nam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	812.6	70.7	926	14	BQ706499
2	781.8	68.0	896	14	BQ672632
3	781	67.9	923	14	BQ950768
4	744.8	64.8	1043	14	BQ072025
5	726.2	63.1	901	12	BG475554
6	682	59.3	988	14	BQ674703

7	628.4	54.6	909	14	BQ646351
8	592	51.5	621	12	BG770181
9	591	51.4	861	12	BG764061
10	577.6	50.2	890	12	BE907778
11	574.8	50.0	784	12	BG763918
12	560.6	48.7	788	12	BG750463
13	551.2	47.9	902	12	BE910036
c 14	544	47.3	618	14	BQ331885
15	529	46.0	774	13	BI760610
16	499	43.4	932	12	BF686586
17	494.4	43.0	496	14	BM705217
18	484.6	42.1	942	12	BG323704
19	475.4	41.3	489	9	AL121548
20	463	40.3	861	12	BG420356
21	457	39.7	1026	13	BM551766
22	454	39.5	454	14	BM728340
23	412.4	35.9	757	10	BE282962
24	398	34.6	719	12	BG248068
c 25	389.2	33.8	498	12	BF955437
26	368.2	32.0	703	13	BJ059899
27	366	31.8	449	10	BB840008
28	362.4	31.5	645	10	BE376581
29	356.6	31.0	861	12	BF538434
30	353	30.7	362	12	BF849713
31	350.2	30.5	565	13	BI663808
c 32	343.8	29.9	358	12	BE812222
33	340.6	29.6	835	12	BG829192
34	340.4	29.6	741	12	BG122769
35	339.4	29.5	436	12	BF955445
36	337.4	29.3	351	14	D81819
37	335.4	29.2	434	10	AW377949
38	331.6	28.8	616	12	BE912488
39	327.6	28.5	424	10	AW377971
40	326.8	28.4	466	14	R14694
41	323	28.1	516	13	BI898950
42	322.6	28.1	475	14	BQ301106
c 43	314.6	27.4	876	14	BQ221203
44	309.2	26.9	736	14	BM963756
45	304	26.4	414	10	BE271866

ALIGNMENTS

RESULT 1
BQ706499
LOCUS BQ706499 926 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8474935 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301634
5', mRNA sequence.
ACCESSION BQ706499.1 GI:21845398
VERSION BQ706499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 926)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCN2517 row: n column: 03
High quality sequence stop: 601.
Location/Qualifiers 1..926
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6301634"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      242 a      247 c      254 g      183 t
ORIGIN
Query Match      70.7%; Score 812.6; DB 14; Length 926;
Best Local Similarity 98.8%; Pred. No. 2.7e-197;
Matches 829; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
QY 269 GACGACGATGAGGATGAGGCGCGGAGGAGAGATTCCTCAACTGATAAAGCGTACTTCAT 328
Db 1 GACGGACGATGAGGATGAGGCGCGGAGGAGATTCCTCAACTGATAAAGCGTACTTCAT 60
QY 329 AGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTC 388
Db 61 AGCTAAGGAAGTGTCTACCAACGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTC 120
QY 389 GTGGTTTCAGACACAGTGAAGCAAGAGGACGCGCATGCGGAGCACTGAAAGTCTCAT 448
Db 121 GTGGTTTCAGACACAGTGAAGCAAGAGGACGCGCATGCGGAGCACTGAAAGTCTCAT 180
QY 449 ATTCCCGAATTTTGAACCTTGCACAAATTCATACATAATTTCTCAAGAAATTCAGCA 508
Db 181 ATTCCCGAATTTTGAACCTTGCACAAATTCATACATAATTTCTCAAGAAATTCAGCA 240
QY 509 ACGACTTGCCCTGTGGAAAGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATCGG 568
Db 241 ACGACTTGCCCTGTGGAAAGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATCGG 300
QY 569 CGATGTCTGTGAAGAACATTCAGGGCATGAAGCACTTGGCGGCTCAGCTGTGGAAGCA 628
Db 301 CGATGTCTGTGAAGAACATTCAGGGCATGAAGCACTTGGCGGCTCAGCTGTGGAAGCA 360
QY 629 CAGCAGGCGCTTGGAGCGCTCGAGATGGAATCGAAGCTCCCGCGGCTGGAGAACTT 688
Db 361 CAGCAGGCGCTTGGAGCGCTCGAGATGGAATCGAAGCTCCCGCGGCTGGAGAACTT 420
QY 689 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTTACCGCTCAACACCTTCTCTCGG 748
Db 421 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTTACCGCTCAACACCTTCTCTCGG 480
QY 749 GCACATGACCGGCTCATGCACTAAGCAGGTCTCGAGCGGCTGTGCAACACCAACCC 808
Db 481 GCACATGCAACCGGCTCATGCACTAAGCAGGTCTCGAGCGGCTGTGCAACACCAACCC 540
QY 809 GCGAGCCAGCGGCTTCAAGGACTTCCGAGCGCTTGGCAGAGATCAGCAGATGCT 868
Db 541 GCGAGCCAGCGGCTTCAAGGACTTCCGAGCGCTTGGCAGAGATCAGCAGATGCT 600
QY 869 GGCACAGCTCCAGGTCAGATGATGAGAAATTTCCAGAACTGCAGAACTCAA 928
Db 601 GGCACAGCTCCAGGTCAGATGATGAGAAATTTCCAGAACTGCAGAACTCAA 660
QY 929 GAAAGATTGATGGCATTCACAAATCTTGGTTCGCGGAAGGAGTTTCATCGTCTGGG 988
Db 661 GAAAGATTGATGGCATTCACAAATCTTGGTTCGCGGAAGGAGTTTCATCGTCTGGG 720
QY 989 CAGCCTCAGCAAGCTCTCGGGAAGGCGCTCCAGACGCGCATGTTCTTCTTCAACCA 1048
Db 721 CAGCCTCAGCAAGCTCTCGGGAAGGCGCTCCAGACGCGCATGTTCTTCTTCAACCA 780
QY 1049 CFTCTGTATACACAG-CCGGGGGCTGACGGCCTTCCAATCAGTTTAAAGTCCACGGG 1106
```

```
Db 781 CGTCTGCTATACAGAGCCGCGGGGTGACGCCCTCAATCAGTTTAAGTCCACGGG 839

RESULT 2
LOCUS      BQ672632               896 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8185043 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6257346
            5', mRNA sequence.
ACCESSION  BQ672632
VERSION    BQ672632.1      GI:21783466
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 896)
AUTHORS   NIH-MGC http://mhc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2411 row: h column: 19
            High quality sequence stop: 630.

FEATURES             source
            1..896
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:6257346"
            /clone_lib="NIH_MGC_102"
            /tissue_type="epidermoid carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
            Site_2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Library constructed
            by Ling Hong in the laboratory of Gerald M. Rubin
            (University of California, Berkeley) using ZAP-cDNA
            synthesis kit (Stratagene) and Superscript II RT (Life
            Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      230 a      257 c      253 g      156 t
ORIGIN
Query Match      68.0%; Score 781.8; DB 14; Length 896;
Best Local Similarity 98.3%; Pred. No. 2.1e-189;
Matches 811; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
QY 15 GCGGACGAGCGCGCTCGGCGCCACGAGGAGGAGGAGTCTTAAAGATAGGACC 74
Db 1 GCGGACGAGCGCGCTCGGCGCCACGAGGAGGAGGAGTCTTAAAGATAGGACC 60
QY 75 CAGCAGAGTAAACCTCAGCCCCCGCAGCAAGCAGAGCTCCCTGACTGGCAGTCTCTAC 134
Db 61 CAGCAGAGTAAACCTCAGCCCCCGCAGCAAGCAGAGTCTCCCTGACTGGCAGTCTCTAC 120
QY 135 CTTTCCGAGCTGTCTGTAACCTCGCAGGGGGGAGTGGCCCTGCCAACGTGACCTTGCT 194
Db 121 CTTTCCGAGCTGTCTGTAACCTCGCAGGGGGGAGTGGCCCTGCCAACGTGACCTTGCT 180
QY 195 CCCAACCTCAGCCCCCGCACACCAAGCAGGCTCTCCTTGTATCAGCCGCTGCTGAATGAC 254
Db 181 CCCAACCTCAGCCCCCGCACACCAAGCAGGCTCTCCTTGTATCAGCCGCTGCTGAATGAC 240
QY 255 CAGGCGTCCCCCGCAGCAGCATGAGGATGAGGCGCGGAGGAAGAGATTCACCACTGAT 314
Db 241 CAGGCGTCCCCCGCAGCAGCATGAGGATGAGGCGCGGAGGAAGAGATTCACCACTGAT 300
```

```
QY 315 AAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCGAGCGAACATATCTCTGAAGGATCTC 374
|||||
Db 301 AAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCGAGCGAACATATCTGAAGGATCTC 360
QY 375 GAAGTTATCACTTCGTGGTTTCAGACACAGTGTAGCAAGAGAGACGCCATGCCGGAAGCA 434
|||||
Db 361 GAAGTTATCACTTCGTGGTTTCAGACACAGTGTAGCAAGAGAGACGCCATGCCGGAAGCA 420
QY 435 CTGAAAGTCTCATATATCCCGAATTTTGAACCTTTTGCACAAATTTTCATACTAATTTTCTC 494
|||||
Db 421 CTGAAAGTCTCATATATCCCGAATTTTGAACCTTTTGCACAAATTTTCATACTAATTTTCTC 480
QY 495 AAGGAATTTGAGCAAGCACTTGGCCCTGTGGGAAGCGCGTCAAAATGCCCAAAATCAGAGAT 554
Db 481 AAGGAATTTGAGCAAGCACTTGGCCCTGTGGGAAGCGCGTCAAAATGCCCAAAATCAGAGAT 540
QY 555 TACCAAGAAATCGGCATGTCATGCTGAAGAACATTTAGGGGCATGAGCACCTGGCGGCT 614
Db 541 TACCAAGAAATCGGCATGTCATGCTGAAGAACATTTAGGGGCATGAGCACCTGGCGGCT 600
QY 615 CACCTGTGGAAGCACAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGG 674
Db 601 CACCTGTGGAAGCACAGCAGGCGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGG 660
QY 675 CGCCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAAC 734
Db 661 CGCCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAAC 720
QY 735 ACCTTCTCTCGGGCCATGTCACCGGCTGATGCACTACAGCAGGCTCTGG-AGCGGCT 793
Db 721 ACCTTCTCTCGGGCCATGTCACCGGCTGATGCACTACAGCAGGCTCTGGAGCGGCT 780
QY 794 GTGCAAAACACACCGCCG-AGCCACGCGACTTCAGGGAGCTGCC 837
Db 781 GTGCAAAACACACCGCCGAGCGCGCAACTTCAGGGAACGGC 825

RESULT 3
BO950768
LOCUS
DEFINITION
AGENCOURT.8842056 Lupski_sciatic_nerve Homo sapiens cDNA clone
IMAGE:6204242 5', mRNA sequence.
BO950768
VERSION
BO950768.1 GI:22366246
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 923)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13625 row: d column: 03
High quality sequence stop: 585.
Location/Qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204242"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
```

```
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGCTCCG-3' and
5'-GACTAGTTCTAGATCCGAGCGGCCCTCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
```

BASE COUNT	238 a	257 c	239 g	189 t
ORIGIN				
Query Match	67.9%	Score 781;	DB 14;	Length 923;
Best Local Similarity	96.0%	Pred. No. 3.4e-189;		
Matches 834;	Conservative 0;	Mismatches 30;	Indels 5;	Gaps 3;
QY 97	CGCAGCCAAAGCACAGGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAAC	156		
Db	31 CTATATCCCAACAGGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAAC	90		
QY 157	CGCAGGGGGAGTGGCCCTGCGCAACGTCGTCGTCCTCCCAACCTTGAGCCCGACACCA	216		
Db	91 CGCAGGGGGAGTGGCCCTGCGCAACGTCGTCGTCCTCCCAACCTTGAGCCCGACACCA	150		
QY 217	AGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACAGGCTGCCCGGACGAGC	276		
Db	151 AGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACAGGCTGCCCGGACGAGC	210		
QY 277	ATGAGATGAGGCGCGGAGGAGAGATTCCTCAACTGATAAAGCTACTTCATAGCTAAG	336		
Db	211 ATGAGATGAGGCGCGGAGGAGAGATTCCTCAACTGATAAAGCTACTTCATAGCTAAG	270		
QY 337	AAGTGTCTACCGAGCGAACATATCTGAAGGATCTGAAGTATTCATCTTCGTGGTTTC	396		
Db	271 AAGTGTCTACCGAGCGAACATATCTGAAGGATCTGAAGTATTCATCTTCGTGGTTTC	330		
QY 397	AGAGCACAGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTC	456		
Db	331 AGAGCACAGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTC	390		
QY 457	ATTTTGAACCTTTTGACAAAATTTTCACTAATAATTTTCTCAAGGAAATTGAGCAACGACTTG	516		
Db	391 ATTTTGAACCTTTTGACAAAATTTTCACTAATAATTTTCTCAAGGAAATTGAGCAACGACTTG	450		
QY 517	CCCTGTGGGAAGCGCCCTCAATGCCAATCAGAGATTACCAAAAGATCGGGGATGTC	576		
Db	451 CCCTGTGGGAAGCGCCCTCAATGCCAATCAGAGATTACCAAAAGATCGGGGATGTC	510		
QY 577	TGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGGCTCACCTGTGGAAGCAGCGAGG	636		
Db	511 TGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGGCTCACCTGTGGAAGCAGCGAGG	570		
QY 637	CCTTGGAGGCCCTTGGAGAATGGAATCAAGAGCTCCCGGGGCTGGAGAATCTTCGAGAG	696		
Db	571 CCTTGGAGGCCCTTGGAGAATGGAATCAAGAGCTCCCGGGGCTGGAGAATCTTCGAGAG	630		
QY 697	ACTTTGAGCTGCAGAAAGTGTGTTACTACCGCTCAACACCTTTCCTTCGCGGCCACTGC	756		
Db	631 ACTTTGAGCTGCAGAAAGTGTGTTACTACCGCTCAACACCTTTCCTTCGCGGCCACTGC	690		
QY 757	ACCGGCTCATGCACACTCAAGCAGGCTCTCGAGCGGCTGTGCAAAACACACCGCCCG	814		
Db	691 ACCGGCTCATGCACACTCAAGCAGGCTCTCGAGCGGCTGTGCAAAACACACCGCCCG	750		
QY 815	CCACGCCGACTTCAGGAGCTGCCGAGCCGC-TTTGGCAGAGATCAC--GGAGATGTGGC	871		
Db	751 CACGCCGACTTCAGGAGCTGCCGAGCCGC-TTTGGCAGAGATCACCGGAAATGGGTGC	810		
QY 872	ACAGCTCCACGGTACGATGATCAAGATGAGAAATTTTCCAGAGCTGCACGAACTCAAGAA	931		

```

Db 811 ACAGCTCCACGGTACGATGAGATGAGAGAAATTTCCCGAAGCTGGCCCAACTCAGAAA 870
QY 932 AGATTTGATTGGCATTGACAATCTTGTGG 960
|||||
Db 871 AGATTTGATTGGCATTGACAATCTTGTGG 899
|||||

RESULT 4
BQ072025 1043 bp mRNA linear EST 02-APR-2002
AGENCOURT_6859787 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5920511
5', mRNA sequence.
ACCESSION BQ072025
VERSION BQ072025.1 GI:19901071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1043)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2105 row: c column: 08
High quality sequence stop: 626.
Location/Qualifiers
1..1043
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 266 a 290 c 294 g 193 t
ORIGIN
Query Match 64.8%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 7.1e-180;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 390 TGGTTTCAGACACAGTGTAGCAAGAGGACGCGATGCCGAGACACTGAAAAGTCTCAT 449
|||||
Db 1 TGGTTTCAGACACAGTGTAGCAAGAGGACGCGATGCCGAGACACTGAAAAGTCTCAT 60
|||||
QY 450 TTCCCGCAATTTTGAACCTTTGCACAATTTTCATACATACTAATTTTCTCAAGAAATTTGAGCA 509
|||||
Db 61 TTCCCGCAATTTTGAACCTTTGCACAATTTTCATACATACTAATTTTCTCAAGAAATTTGAGCA 120
|||||
QY 510 CGACTTGCCTGTGGAGGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATCGGC 569
|||||
Db 121 CGACTTGCCTGTGGAGGCGCGCTCAATGCCAAATCAGAGATTACCAAGAAATCGGC 180
|||||
QY 570 GATGTCATCTGAAGACATTCAGGCGCATGAAGACACCTGGCGGCTCACCTGTGGAGCAC 629
|||||
Db 181 GATGTCATCTGAAGAACATTCAGGCGCATGAAGACACCTGGCGGCTCACCTGTGGAGCAC 240
|||||

```

```

QY 630 AGCGAGGCTTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 689
|||||
Db 241 AGCGAGGCTTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 300
|||||
QY 690 TGCAGAGACTTTTGTAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGG 749
|||||
Db 301 TGCAGAGACTTTTGTAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGGG 360
|||||
QY 750 CCAGTGCACCGGCTCATGCACACTACAAGCAGGTCCTGGAGCGGCTGTGCAAAACACACCGG 809
|||||
Db 361 CCAGTGCACCGGCTCATGCACACTACAAGCAGGTCCTGGAGCGGCTGTGCAAAACACACCGG 420
|||||
QY 810 CCGAGCCACCGCGCACTTCAGGCACTGCCAGCGCTTTGGCAGAGATCACGGAGATGGTG 869
|||||
Db 421 CCGAGCCACCGCGCACTTCAGGCACTGCCAGCGCTTTGGCAGAGATCACGGAGATGGTG 480
|||||
QY 870 GCACAGCTCCACGGTACGATGATCAAGATGAGAGATTTCAGAAAGCTGCACAACTCAAG 929
|||||
Db 481 GCACAGCTCCACGGTACGATGATCAAGATGAGAGATTTCAGAAAGCTGCACAACTCAAG 540
|||||
QY 930 AAAGATTTGATTGGCATTCACAATCTTGTGGTTCGGGGAAGGAGTTTCATCGTCTGGGC 989
|||||
Db 541 AAAGATTTGATTGGCATTCACAATCTTGTGGTTCGGGGAAGGAGTTTCATCGTCTGGGC 600
|||||
QY 990 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACAGAC 1049
|||||
Db 601 AGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACAGAC 660
|||||
QY 1050 GTCTGTCTATACAGCGCGGGGCTGAGCGGCTCCAAATCAGTTTAAAGTCACCGGGCAG 1109
|||||
Db 661 GTCTGTCTATACAGCGCGGGGCTGAGCGGCTCCCAATCAGTTTAAAGTCACCGGGCAG 720
|||||
QY 1110 CTCGCGCTCTATGGCATGACGATTGAGG 1137
|||||
Db 721 CTCGCGCTCTATGGCATGACGATTGAGG 748
|||||

RESULT 5
BQ475554 901 bp mRNA linear EST 21-MAR-2001
LOCUS 602491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
mRNA sequence.
ACCESSION BQ475554
VERSION BQ475554.1 GI:13407833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTp
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1381 row: h column: 11
High quality sequence stop: 818.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4620346"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
FEATURES
source

```


/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 232 a 243 c 253 g 173 t

ORIGIN

Query Match 63.1%; Score 726.2; DB 12; Length 901;
Best Local Similarity 91.3%; Pred. No. 3.8e-175;
Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;

QY 229 CTTTGATCAGCCGCTGCTGAATGACAGGCGCTGCCCGGACGCGATGAGGATGAGG 288
|||||
Db 2 CTTTGATCAGCCGCTGCTGAATGACAGGCGCTGCCCGGACGCGATGAGGATGAGG 61
|||||

QY 289 GCCGGAGGAGAGATTCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTGTCTACCA 348
|||||
Db 62 GCCGGAGGAGAGATTCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTGTCTACCA 121
|||||

QY 349 CCAGGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGAGTGA 408
|||||
Db 122 CCAGGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGAGTGA 181
|||||

QY 409 GCAAAGAGGAGCGCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTTGAACCTT 468
|||||
Db 182 GCAAAGAGGAGCGCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTTGAACCTT 241
|||||

QY 469 TGCACAAATTTTACTATAATTTTCTCAAGGAATTTGAGCAACGACTTGCCCTGTGGGAAG 528
|||||
Db 242 TGCACAAATTTTACTATAATTTTCTCAAGGAATTTGAGCAACGACTTGCCCTGTGGGAAG 301
|||||

QY 529 GCCGCTCAATGCCCAATCAGAGATTACCAAGATTCGGCGATGTCATGCTGAAGAACA 588
|||||
Db 302 GCCGCTCAATGCCCAATCAGAGATTACCAAGATTCGGCGATGTCATGCTGAAGAACA 361
|||||

QY 589 TTCAGGCGATGAAGCACTTGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTCGAGGCC 648
|||||
Db 362 TTCAGGCGATGAAGCACTTGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTCGAGGCC 421
|||||

QY 649 TGGAGATGGAATCAAGAGTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGC 708
|||||
Db 422 TGGAGATGGAATCAAGAGTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGC 481
|||||

QY 709 AGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCACTGCACCGGCTCATGC 768
|||||
Db 482 AGAGGTGTGTACCTACCGCTCAACACCTTCTCTCTGGGGCCACTGCACCGGCTCATGC 541
|||||

QY 769 ACTACAAGCAGGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCA 828
|||||
Db 542 ACTACAAGCAGGTCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCA 601
|||||

QY 829 -GGGACTGCGGAGCCGCTTTGGCAGAGATCACGGAGATGGTGGCAGAGTCCACGGTACG 887
|||||
Db 602 GGGGACTGGGAGCCGCTTTGGCAGAGATCACGGAGATGGTGGCAGAGTCCACGGTACG 661
|||||

QY 888 ATGATCAAGATGGAGATTTCCAGAGCTGCAGCACTCAAGAAAGATTGATGGCAAT 947
|||||
Db 662 ATGATCAGATGGAGATTTCCAGAGCTGCCGAATCTCAAGACAGATTTGATGGCAAT 721
|||||

QY 948 GACAATCTGTGTCTCCGGAA- - - -GGAGTTTCATCCGCTCTGGGCGGCTCAGCAA- - - 1000
|||||

Db 722 GACAATCTGTGTGGTCCCGGGAAGCGGAGTTCCATCCGGTCTGGGCGGCTTCAAGCAAGC 781
|||||

QY 1001 -GCTCTCGGGGAGGGGCTCCAGCAGCGGATGTTCTTCTCTGTTCAACAGAGCTCCT-GCTA 1058
|||||

Db 782 TCCTCGGGGGAAGGACTCCCAAGCAGCGCATGCTCTCTCTGTTCAAGAGAGTCTCGCTT 841
|||||

QY 1059 TACACAGCGGGGGTGTGACGGCCTCCATACAGTTTAAAGTCCACGGGCGAGTCTC 1113
|||||

Db 842 TACACAGCGCGGCTGACAGGCGTCCAATCCAGCTTAAGCAAGGCGAGTCCC 896

RESULT 6
BQ674703 988 bp mRNA linear EST 15-JUL-2002

LOCUS AGENCOURT_818989 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255217

DEFINITION 5', mRNA sequence.

ACCESSION BQ674703

VERSION BQ674703.1 GI:21785537

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 988)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLM2405 row: p column: 02
High quality sequence stop: 685.

FEATURES
Location/Qualifiers
1..988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6255217"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 216 a 306 c 291 g 161 t 14 others

ORIGIN

Query Match 59.3%; Score 682; DB 14; Length 988;
Best Local Similarity 99.6%; Pred. No. 8.6e-164;
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 456 AATTTTGAACCTTTGACAAATTTTCTCAAGGAAATTTGAGCAACGACTT 515
|||||

Db 1 AATTTTGAACCTTTGACAAATTTTCTCAAGGAAATTTGAGCAACGACTT 60
|||||

QY 516 GCCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGC 575
|||||

Db 61 GCCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGC 120
|||||

QY 576 ATGCTGAAGAAACATTCAGGCGCATGAAGCACCTTGGCGGCTCACCTGTGGAAGCACACGAG 635
|||||

Db 121 ATGCTGAAGAAACATTCAGGCGCATGAAGCACCTTGGCGGCTCACCTGTGGAAGCACACGAG 180
|||||

QY 636 GCCTTGGAGGCCCTGGAGAAATCAAGAGCTCCCGGCGCTGGAGAACTTCTCAGA 695
|||||

Db 181 GCCTTGGAGGCCCTGGAGAAATCAAGAGCTCCCGGCGCTGGAGAACTTCTCAGA 240
|||||

QY 696 GACTTTGAGCTGCAGAGGTTGTTACCTACCGCTCAACACCTTCTCTCTCGGGCCACTG 755
|||||

Db 241 GACTTTGAGCTGCAGAGGTTGTTACCTACCGCTCAACACCTTCTCTCTCGGGCCACTG 300
|||||

```
QY 756 CACCGGCTCATGCTACAAAGAGGTCTCGAGCGGCTGTGCAAAACACACCCGCGGAGC 815
Db 301 CACCGGCTCATGCTACAAAGAGGTCTCGAGCGGCTGTGCAAAACACACCCGCGGAGC 360
QY 816 CAGCCGACTTCAGGACTGCGGAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAC 875
Db 361 CAGCCGACTTCAGGACTGCGGAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAC 420
QY 876 CTCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCACGAATCTCAAGAAAGAT 935
Db 421 CTCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCACGAATCTCAAGAAAGAT 480
QY 936 TTGATTTGGCATTTGACAAATCTTGTTGGTTCCGGAAGGAGTTTCATCGTCTGGGACGCTTC 995
Db 481 TTGATTTGGCATTTGACAAATCTTGTTGGTTCCGGAAGGAGTTTCATCGTCTGGGACGCTTC 540
QY 996 AGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACAGCTCCCTG 1055
Db 541 AGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACAGCTCCCTG 600
QY 1056 CTATACACGAG-CCGGGGGCTGACGGCTCCCAATCAAGTTTAAAGTCCACGGCAGCTCCC 1114
Db 601 CTATACACGAGNCCGGGGGCTGACGGCTCCCAATCAAGTTTAAAGTCCACGGCAGCTCCC 560
QY 1115 GCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 1150
Db 661 GCTCTATGGCATGACGATTGANGAGAGCGAAGACNA 696

RESULT 7
BQ646351
LOCUS
DEFINITION BQ646351 909 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8286354 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299462
5', mRNA sequence.
ACCESSION BQ646351
VERSION BQ646351.1 GI:21770523
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM2512 row: c column: 15
High quality sequence stop: 589.
Location/Qualifiers
1. 909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6299462"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="pDHL0B (phage-resistant)"
/note="organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
```

FEATURES
Source

```
BASE COUNT 179 a 332 c 235 g 163 t
ORIGIN
Query Match 54.6%; Score 628.4; DB 14; Length 909;
Best Local Similarity 97.6%; Pred. No. 4.5e-150;
Matches 638; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 497 GGAATTTGAGCAACGACTTTCCTGTTGGGAAGCGGCTCAAAATGCCAAATCAGAGATT 556
Db 1 GGAATTTGAGCAACGACTTTCCTGTTGGGAAGCGGCTCAAAATGCCAAATCAGAGATT 60
QY 557 CCAAGAATTCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGACCTGGCGGCTCA 616
Db 61 CCAAGAATTCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGACCTGGCGGCTCA 120
QY 617 CTTGTGGAGACACGCGAGGCTTGGAGGCCCTGGAGAAATGAATCAAGAGCTCCCGGCG 676
Db 121 CTTGTGGAGACACGCGAGGCTTGGAGGCCCTGGAGAAATGAATCAAGAGCTCCCGGCG 180
QY 677 GCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC 736
Db 181 GCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC 240
QY 737 CTTCTCTCTGCGGCCACTGCACCGGCTCATGACACTACAAAGAGTCTCTGGAGCGGCTGTG 796
Db 241 CTTCTCTCTGCGGCCACTGCACCGGCTCATGACACTACAAAGAGTCTCTGGAGCGGCTGTG 300
QY 797 CAACACACACCGCGGCGGCGGCGGCTTTCAGGAGCTCCGAGCGGCTTTCGAGAGAT 856
Db 301 CAACACACACCGCGGCGGCGGCGGCTTTCAGGAGCTCCGAGCGGCTTTCGAGAGAT 360
QY 857 CACGAGATGTTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAAAGCT 916
Db 361 CACGAGATGTTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAAAGCT 420
QY 917 GCACGAACCTCAAGAAGATTTGATTGGCAATTCACAAATCTTGTGGTTCGCGGAGGAGTT 976
Db 421 GCACGAACCTCAAGAAGATTTGATTGGCAATTCACAAATCTTGTGGTTCGCGGAGGAGTT 480
QY 977 CATCGCTGCGGCGGCGGCGGCGGCTTCGCGGAGGCGGCTCCAGCAGCGATGTTCTTT 1036
Db 481 CATCGCTGCGGCGGCGGCGGCGGCTTCGCGGAGGCGGCTCCAGCAGCGATGTTCTTT 540
QY 1037 CTTGTTCAACGAGCTCTCTGCTATACACGAGCGGCGGCTTCGCGGAGGCGGCTTCGTTAA 1096
Db 541 CTTGTTCAACGAGCTCTCTGCTATACACGAGCGGCGGCTTCGCGGAGGCGGCTTCGTTAA 600
QY 1097 AGTCCACGGGCGAGTCCCGCTCTATGGCATGACGATTGAGGAGCGAAGACGA 1150
Db 601 AGTCCACGGGCGAGTCCCGCTCTATGGCATGACGATTGAGGAGCGAAGACGA 654

RESULT 8
BQ770181
LOCUS
DEFINITION BQ770181 621 bp mRNA linear EST 15-MAY-2001
602744940F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4878177 5',
mRNA sequence.
ACCESSION BQ770181
VERSION BQ770181.1 GI:14080834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCFD/PTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1763 row: 0 column: 10
High quality sequence stop: 621.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

FEATURES	SOURCE
1. .621	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:4878177"	
/clone_lib="NIH_MGC_49"	
/tissue_type="melanotic melanoma, high MDR (cell line)"	
/lab_host="Drl10B (phage-resistant)"	
/note="Organ: Skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGACG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. "	
144 a	189 q 115 t
BASE COUNT	

```

Library: 1
BASE COUNT      144 a      173 c      189 g      115 t
ORIGIN

Query Match      51.5%; Score 592; DB 12; Length 621;
Best Local Similarity 100.0%; Pred. No. 7.5e-141;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 559 AAAGAATCGCGGATGTCATGCTGAAGAACATTTCAGGGCATGAAGCACCCTGGCGGGTCACC 618
|||||
Db 2 AAAGAATCGCGGATGTCATGCTGAAGAACATTTCAGGGCATGAAGCACCCTGGCGGGTCACC 61

QY 619 TGTGGAGCACACGAGGCGCTTGGAGGCCCTGGAGAAATGGAATCACAGACTCCCGCGCGC 678

Db	62	TGTGGAAGCACACGAGGCTTGGAGGCCCTGGAGAAATGGAATCAAGAGCTCCCGGCGGC	121
QY	679	TGGAGAACTTCGACAGAGACTTTGAGCTGCAGAGGTTGTTTACCTACCGCTCAACACCT	738

Db
122 TGGAGAACTTCACAGAGACTTTGAGCTGCAGAGGTTGTTACCTACCGCTCAACACCT 181

	Qy	Db
739	TCCTCCTGCGGCACTGCACCGGCTCATGCATACAAAGCAGGTCTCTGGAGCGGCTGTGCA	798
182	TCCTCCTGCGGCACTGCACCGGCTCATGCATACAAAGCAGGTCTCTGGAGCGGCTGTGCA	241

Qy 799 AACACACCCGGCGAGCCGACGGCGACTTCAGGACTGCCAGCGCGTTGGCAGAGATCA 858

Db	242	AACACACCCGGCAGCCAGCCCGACTTCAGGGA	CTGCCGAGCCGCTTGGCAGAGATCA	301
Qv	859	CGGAGTGGTGGCACAGCTCCACCGTACGATGATCA	AAGATGGAGAAATTCACGAAGCTGC	918

Db 302 CGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGGAGAAATTCAGAGAAGCTGC 361

Qy	919	ACGA	ACTCAAG	AAAAGATT	TTCAT	TGGCA	TTGACA	ATCTT	TGGT	TCCGG	GAAGG	AGATTCA	978
Db	362	ACGA	ACTCAAG	AAAAGATT	TTCAT	TGGCA	TTGACA	ATCTT	TGGT	TCCGG	GAAGG	AGATTCA	421

Qy 979 TCCGTCGGCAGCCTCAGCAAGCTCTCGGGGAGGGGCTCCAGCAGGCATGTTCTTCC 103

Db	422	TCCGTCGGGACGCTCAGCAAGCTCTCGGGGAGGGGCTCCAGCAGCGCATGTTCTTCC	481
Oy	1039	TGTTCAACGAGCTCTGCTATACAGACCGGGGGCTGACGGCTCCCAATCAGTTTAAAG	109

Db 482 TGTTCACGACGCTCTGCTATACACAGCGCGGGCTGACGGCCTCCAATCAGTTAAAG 541

QY 1099 TCCACGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAGCGGAACGCA 1150
|||||
543 TCCACGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAGCGGAACGCA 503

DECEMBER 0 1942

RESULTS ?
BG764061

LOCUS	861 bp	mRNA	linear	EST 15-MAY-2001
DEFINITION	60273707f1	NIH_MGC_49	Homo sapiens cDNA clone IMAGE:4862417 5',	

Accession	Version	miRNA sequence
BG764061		
BG764061.1	GI:14074714	

ORGANISM Homo sapiens
SOURCE human.
EST. SOURCE
KEYWORDS

REFERENCE

ALLAN, R. D., AND J. A. ALLAN. 1968. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

BLOOM, F. S. 1970. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

CAMP, C. E. 1972. Mammals as a monophyletic group. *J. Mammal.* 53:1-10.

DODD, R. K. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

HALL, B. L. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

KRÖGER, H. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

LILLIE, R. W. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

MILNE, G. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

NELSON, R. E. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

PETERSON, T. J. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

RIDGWAY, R. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

SIMPSON, G. G. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

TAYLOR, W. H. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

WATKINS, P. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

ZIMMERMAN, J. R. 1972. The evolution of the placental mammals. Univ. Chicago Press, Chicago.

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT:
Contact: Robert Strausberg, Ph.D.
Email: ccapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

FEATURES
Location/Qualifiers
High quality sequence stop: 817.
plate: L1CMI722 row: n column: 18

```
source
1.861
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/clone="IMAGE:4862417"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"

```

```

/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC

BASE COUNT	198 a	245 c	270 g	147 t	1 others
ORIGIN	Library. "				

Query Match 51.4%; Score 591; DB 12; Length 861;
Best Local Similarity 99.8%; Pred. NO. 1.7e-140;

	Matches	591;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
QY	559	AAAGAATCGGGGATGCTGCTGAAGAACATTCAGGGCATGAAGACACCTGGCGGCTCAC	618							

db 2 AAAGAATCGGCATGCTATGTAAGAACATTGAGGCATGAAGCACCTGGCGGCTACC 61

Qy	619	TTGTGGAAACACACGAGCGCCCTTGGAGSCCTCGAGATGGAATCAAGAGCTCCCGGCGC	678
Db	62	TTGTGGAAGCACACGAGCGCCCTTGGAGSCCTCGAGATGGAATCAAGAGCTCCCGGCGC	121

QY 679 TGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAGGTGTATTACCTACCGGCTCAACACCT 738

Db	122	TGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCT	181
Ov	739	TCCTCTCCGGGCACACTGCACCGGCTCATGTCACTACAGCAGGTCTCTGGAGCGGCTGTGCA	798

db 182 TCCTCCTGCGGCACATGCACCGGCTCATGCTACAGCAGGTCCCTGGAGCGGCTGTGCA 241

Qy	799	ARACCAACCGCGAGCCACGCCGACTTCAGGAGATGCCGAGCCGCTTTGGCAGAGATCA	858
Dh	242	ARACCAACCGCGAGCGACGCGGACTTCAGGAGATGCCGAGCCGCTNTGGCAGAGATCA	301

859 CGGAGATGGTGGCACAGCTCCACGGTACGATCATCAAGATGGAGAATTTCAGAGAAGCTGC 918

Db 302 CGGAGATGGTGGCACAGCTCCACGGTACGATCAAGATGGAGAATTTCCAGAACTGC 361

QY 919 ACGAAGCTCAAGAAAGATTGATGGCATGTGACAATCTTGTGTGTTCCGGGAAGGAGTTCA 978
 |||||||
 Db 362 ACGAAGCTCAAGAAAGATTGATGGCATGTGACAATCTTGTGTGTTCCGGGAAGGAGTTCA 421
 |||||||
 QY 979 TCCGTTCTGGCAGCTCAGCAGCTCTCGGGAAGGGCTCCAGCAGGCGATGTTCTTCC 1038
 |||||||
 Db 422 TCCGTTCTGGCAGCTCAGCAGCTCTCGGGAAGGGCTCCAGCAGGCGATGTTCTTCC 481
 |||||||
 QY 1039 TGTTCACACAGCTCTGCTATACAGAGCGGGGCTGACGGCCTCCAATCAGTTTAAAG 1098
 |||||||
 Db 482 TGTTCACACAGCTCTGCTATACAGAGCGGGGCTGACGGCCTCCAATCAGTTTAAAG 541
 |||||||
 QY 1099 TCCAGGGCAGCTCCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 1150
 |||||||
 Db 542 TCCAGGGCAGCTCCCGCTCTATGTCATGACGATTGAGGAGCGAAGACGA 593
 |||||||

RESULT 10
 BE907778 890 bp mRNA linear EST 20-OCT-2000
 LOCUS 601502061F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903855 5',
 mRNA sequence.

ACCESSION BE907778
 VERSION BE907778.1 GI:10401682
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 890)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9708 row: j column: 16
 High quality sequence stop: 646.

FEATURES
 1..890
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3903855"
 /clone_lib="NIH_MGC_70"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 BASE COUNT 251 a 242 c 241 g 156 t
 ORIGIN

Query Match 50.2%; Score 577.6; DB 12; Length 890;
 Best Local Similarity 98.4%; Pred. No. 4.7e-137;
 Matches 657; Conservative 0; Mismatches 4; Indels 7; Gaps 7;

QY 485 TAATTTCTCAAGGAATGAGCAAGCTTCCTGCTGGGAGGCGCTCAATGCCCA 544
 |||||||
 Db 1 TAATTTCTCAAGGAATGAGCAAGCTTCCTGCTGGGAGGCGCTCAATGCCCA 60
 |||||||

QY 545 AATCAGAGATTACCAAGAAATCGGCATGTCATGCTGAAGACATTCAGGGCATGAAGCA 604
 |||||||
 Db 61 AATCAGAGATTACCAAGAAATCGGCATGTCATGCTGAAGACATTCAGGGCATGAAGCA 120
 |||||||

QY 605 CTGG-CGGCTCAGCTGTGGGAAGCAGCGAGGCTTGGAGGCGCTTGGAGATGAATCA 663
 |||||||

Db 121 CTGGAGCGCTCACTCTGTGAAGCACACGAGGCGCTTGGAGGCGCTTGAGAAATGGAATCA 180
 |||||||
 QY 664 AGAGTCTCCGGCGGC-TGGAGAACTTC-TGCAGAGACTTTGAGCTGCGAGAGGTTGTTA 721
 |||||||
 Db 181 AGAGTCTCCGGCGGCATGGAGAACTTCATGACAGACTTTGAGCTGCGAGAGGTTGTTA 240
 |||||||
 QY 722 CCTACCG-CTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGG 780
 |||||||
 Db 241 CCTACCGACTCAACACCTTCTCTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGG 300
 |||||||
 QY 781 TCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCAGGGACTCCCGAG 840
 |||||||
 Db 301 TCTTGGAGCGGCTGTGCAAAACACACCGCGGAGCCAGCGGACTTCAGGGACTCCCGAG 360
 |||||||
 QY 841 CCGCTTTGGCAGAGATCACGAGATGTGGCACAGCTCCACGGTACGATGATCAAGATGG 900
 |||||||
 Db 361 CCGCTTTGGCAGATCACGAGATGTGGCACAGCTCCACGGTACGATGATCAAGATGG 420
 |||||||
 QY 901 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTGATTGGCAATTGACAATCTTGTGG 960
 |||||||
 Db 421 AGAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTGATTGGCAATTGACAATCTTGTGG 480
 |||||||
 QY 961 TTCGGGAAGGAGTTATCGCTGTGGCAGCTCAGCAAGCTCTCGGGGAGGAGGCTCC 1020
 |||||||
 Db 481 TTCGGGAAGGAGTTATCGCTGTGGCAGCTCAGCAAGCTCTCGGGGAGGAGGCTCC 540
 |||||||

QY 1021 AGCAGCGCATGTTCTCTGTTTCAA-CGACGTCTCTGCTATACAGGCGCGGG-CTGAC 1078
 |||||||
 Db 541 AGCAGCGCATGTTCTCTGTTTCAA-CGACGTCTCTGCTATACAGGCGCGGGCTGAC 600
 |||||||

QY 1079 GGCCTCCCAATCAGTTTAA-AGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGG 1137
 |||||||
 Db 601 GGCCTCCCAATCAGTTTAA-AGTCCACGGGCGAGCTCCCGCTCTATGGCATGACGATTGAGG 660
 |||||||

QY 1138 AGAGCGAA 1145
 |||||||
 Db 661 AGAGCGAA 668
 |||||||

RESULT 11
 BE763918 784 bp mRNA linear EST 15-MAY-2001
 LOCUS 602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862444 5',
 mRNA sequence.

ACCESSION BE763918
 VERSION BE763918.1 GI:14074571
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 784)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM1722 row: o column: 21
 High quality sequence stop: 761.

FEATURES
 1..784
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4862444"
 /clone_lib="NIH_MGC_49"
 /tissue_type="melanotic melanoma, high MDR (cell line)"

/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. I"

```
BASE COUNT      203 a      223 c      223 g      135 t
ORIGIN
Query Match      50.0%; Score 574.8; DB 12; Length 784;
Best Local Similarity 99.3%; Pred. No. 2.2e-136;
Matches 598; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CCGGTAAACAGCAGCGGAGCGCGCTCGCGCCACAGGAGGAGGAGGTGCG 60
Db 183 CGGGTAAACAGCAGCGGAGCGCGCTCGCGCCACAGGAGGAGGAGGTGCG 242
QY 61 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCAGCAAGCAGAGCTCCCTGA 120
Db 243 TTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCCAGCAAGCAGAGCTCCCTGA 302
QY 121 CTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGGGGAGTGGCCCTGCCA 180
Db 303 CTGGCAGTCTCACCTTTCCGAGCTGTCTGTGAACCTCGCAGGGGGAGTGGCCCTGCCA 362
QY 181 AGTGACCTTGTCTCCACCTTGGCCCGGACCAAGCAGGCGCTCTCCCTGTATCAGCC 240
Db 363 AGTGACCTTGTCTCCACCTTGGCCCGGACCAAGCAGGCGCTCTCCCTGTATCAGCC 422
QY 241 CGCTGCTGAATGACCAAGCGCTGCCCCCGCAGCAGCATGAGTGAAGCGCGGAGGA 300
Db 423 CGCTGCTGAATGACCAAGCGCTGCCCCCGCAGCAGCATGAGTGAAGCGCGGAGGA 482
QY 301 GATTCCTCACTGATTAAGCGTACTTCTATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 360
Db 483 GATTCCTCACTGATTAAGCGTACTTCTATAGCTAAGGAAGTGTCTACCCAGCGCAACAT 542
QY 361 ATCTGAGATCTCGAAGTATCATCTTCTGTTTTCAGAGCAGTGAAGCAAGAGGAGC 420
Db 543 ATCTGAGATCTCGAAGTATCATCTTCTGTTTTCAGAGCAGTGAAGCAAGAGGAGC 602
QY 421 CCATGCCGGAAGCACTGAAAGTCTCATATTCGCGAATTTTCAACCTTTGCAACAATTC 480
Db 603 CCATGCCGGAAGCACTGAAAGTCTCATATTCGCGAATTTTGAACCTTTGCAACAATTC 662
QY 481 ATACTAATTTTCTCAAGGAAA-TTGAGCAACGACTTGCCCTGTGGGAAGCGCGCTCAAA 539
Db 663 ATACTAATTTTCTCAAGGAAA-TTGAGCAACGACTTGCCCTGTGGGAAGCGCGCTCAAA 722
QY 540 GCCCAATCAGAGATTACCAAGATTCGCGGATGTCATGC-TGAGAAACATTCAGGGCAT 598
Db 723 GCCCAATCAGAGATTACCAAGATTCGCGGATGTCATGC-TGAGAAACATTCAGGGCAT 782
QY 599 GA 600
Db 783 GA 784
```

```
RESULT 12
LOCUS BG750463
DEFINITION 602709323F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4845892 5',
mRNA sequence.
ACCESSION BG750463
VERSION BG750463.1 GI:14061116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 788)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1685 row: n column: 05
High quality sequence stop: 756.
FEATURES
source
location/Qualifiers
1..788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4845892"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. I"

```
BASE COUNT      196 a      210 c      229 g      153 t
ORIGIN
Query Match      48.7%; Score 560.6; DB 12; Length 788;
Best Local Similarity 94.7%; Pred. No. 9.8e-133;
Matches 721; Conservative 0; Mismatches 24; Indels 16; Gaps 13;

QY 405 GTGACCAAGAGAGCGCCATCCGGAAGCAGCTGTAACCTCTCATATTCCTCCGAATTTCA 463
Db 1 GTGACCAAGAGAGCGCCATCCGGAAGCAGCTGTAACCTCTCATATTCCTCCGAATTTCA 60
QY 464 ACCTTTGCACAAA-TTTCATACTAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCCTGT 522
Db 61 ACCTTTGCACAAA-TTTCATACTAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCCTGT 120
QY 523 GGAAGAGCGCGCTCAA-ATGCCCAATTCAGAGATTACCAAGAAATCGGC-GATGTCTATGCT 580
Db 121 GGAAGAGCGCGCTCAA-ATGCCCAATTCAGAGATTACCAAGAAATCGGC-TGATGTCTATGCT 180
QY 581 GAAGAACATTCAGGCGCATGAAGCACCCTGCGGCTCACCTGTGGAAGCAGGAGGCGCTT 640
Db 181 GAAGAACATTCAGGCGCATGAAGCACCCTGCGGCTCACCTGTGGAAGCAGGAGGCGCTT 240
QY 641 GGAGGCGCTTGGAGATGGAATCA-AGAGCTCCCGCGGCTGGAGAACTTCTTCAGAGACT 699
Db 241 GGAGGCGCTTGGAGATGGAATGGAATCAGAGAGCTCCGCGGCTGGAGAACTTCTTCAGAGACT 300
QY 700 TTGAGCTGCAGAAAGTGTGTTACCTA--CCGCTCAACACCTTCTCCTCGGCCCACTGCA 757
Db 301 TTGAGCTGCAGAAAGTGTGTTACCTAGCGGCTCAAGCACCCTTCTCCTCGGCCCACTGCA 360
QY 758 CCGGCTCATGCACTA-CAAGCAGGTCTCTGGAGCGGC--TGTCGAACACACACCGCCGCA 813
Db 361 CCGGCTCATGCACTA-CAAGCAGGTCTCTGGAGCGGCCTGGAGCGGCCTGGAGCGGCCTG 420
QY 814 GCC-ACGCGCACTTCAGGAGCTGCCAGCGCTTTGGCAGAGATCAGGAGATGTGGCA 872
Db 421 GCCAGCGCGCACTTCAGGAGCTGCCAGCGCTTTGGCAGAGATCAGGAGATGTGGCA 480
QY 873 CAGCTCCAGCGTACGATGATCAAGATGGAAGATTTCCAGAACTGCAGCACTCAAGAAA 932
Db 932 CAGCTCCAGCGTACGATGATGGAAGATTTCCAGAACTGCAGCACTCAAGAAA 932
```

```

Db 481 CAGATCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGAACTCAGGAAA 540
QY 933 GATTTGATGGCATTCACAAATCTTGTTCCGG-GAAGGGAGTTTCATCCGCTCGGCGAG 991
Db 541 GATTCGATGGCATTCACAAATCTTGTTCCGGGAAGGGAGTTTCATCCGCTCGGCGAG 600
QY 992 CCTCAGCAAGCTCTCGGGGAA-GGGGCTCCAGCAG-CGCATGTTCTTCTGTTCAACGAC 1049
Db 601 ACTCAGCAAGCTCTCGGGGAACGGGCTCCAGCAGCCGCGATCTTCTGTTCAACGAC 660
QY 1050 GTCCCTGCTATACAGAGCCGGGGCTGACGGGCTCAATCAGTTTAAAGTCACCGGCGAG 1109
Db 661 GTCCTGCTATACAGAGCCGGGGCTGACGG-CTCCAATCAGTTTAAAGTCACCGGCGAG 719
QY 1110 CTCGCCCTCTATGGATGATGAGTTGAGGAGCGGAGAGCA 1150
Db 720 CTCGCCATCTATGGATGATGAGTTGAGGAGCGGAGAGCA 760

RESULT 13
LOCUS BE910036 902 bp mRNA linear EST 20-OCT-2000
DEFINITION 60149808f1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
mRNA sequence.
ACCESSION BE910036
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9698 row: j column: 22
High quality sequence stop: 750.
FEATURES
source
Location/Qualifiers
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN
Query Match 47.9%; Score 551.2; DB 12; Length 902;
Best Local Similarity 94.9%; Pred. No. 2.8e-130;
Matches 592; Conservative 0; Mismatches 28; Indels 4; Gaps 2;

QY 1 CGGGTAACAGCAGCGGACGAGCGGCTCGGCCGCCACGAGGAGGAGGAGGTCG 60
Db 161 CGGGTAACAGCAGCGGACGAGCGGCTCGGCCGCCACGAGGAGGAGGAGGTCG 220
QY 61 TTAAGGATAGGACCCAGCAGAGTAACTCAGCCCCCGCAGCAGCAGCAGCTCCCTCA 120
Db 221 TTAAGGATAGGACCCAGCAGAGTAACTCAGCCCCCGCAGCAGCAGCAGCTCCCTCA 280

```

```

QY 121 CTGGCAGTCTCACCTTTCCGAGCTGTGTGAACTCGCAGGGGAGTGGCCCCCTGCCA 180
Db 281 CTGGCAGTCTCACCTTTCCGAGCTGTGTGAACTCGCAGGGGAGTGGCCCCCTGCCA 340
QY 181 ACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGAGGCGCTCTCCCTTGATCAGCC 240
Db 341 ACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGAGGCGCTCTCCCTTGATCAGCC 400
QY 241 CGCTGCTCAATGACACAGCGCTGCCCGCGAGCGAGGATGAGGATGAGGCGGAGGAAGA 300
Db 401 CGCTGCTCAATGACACAGCGCTGCCCGCGAGCGAGGATGAGGATGAGGCGGAGGAAGA 460
QY 301 GATTTCCCAACTGATAAAGCGTACTTTCATAGCTAAGAAAGTGTCTACCAACCGAGCGAAT 360
Db 461 GATTTCCCAACTGATAAAGCGTACTTTCATAGCTAAGAAAGTGTCTACCAACCGAGCGAAT 520
QY 361 ATCTGAAGGATCTCGAAGTTTATCACTTCGTGGTTTCAGAGCACAGTGAAGAGGAGCG 420
Db 521 ATCTGAAGGATCTCGAAGTTTATCACTTCGTGGTTTCAGAGCACAGTGAAGAGGAGCG 580
QY 421 CCATGCCGG-AAGCACTGAAAA---GTCTCATATATCCCGAATTTTGAACCTTTGCACAAA 476
Db 581 CCATGCCGGAAAGCACTGAAAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAA 640
QY 477 TTTTCATATAATTTTCTCAAGGAAATTCAGCAACGACTTGCCTCTGGGAAGGCCGCTCA 536
Db 641 TTTTCATATAATTTTCTCAAGGAAATTCAGCAACGACTTGCCTCTGGGAAGGCCGCTCA 700
QY 537 AATGCCCAAACTCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAATTCAGGGC 596
Db 701 AATGCCCAAACTCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAATTCAGGGC 760
QY 597 ATGAGCAGCTTGGCGCTCACCTG 620
Db 761 ATAAAGCACCGGGGGTCCCTGTG 784

RESULT 14
LOCUS BQ331885/6 618 bp mRNA linear EST 17-MAY-2002
DEFINITION PM3-ET0268-100501-005-d11 ET0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ331885
VERSION BQ331885.1 GI:20973153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3st2-PM3-ET0268-
100501-005-d11&t3=2001-05-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 572.

```


Db 726 ATACTAATTTCTCAAGGAAATTGAGCAACGACTTCCCTGTGGGAAAGG 774
|||||

Search completed: December 3, 2002, 21:25:20
Job time : 1257.16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:35:00 ; Search time 4141.44 Seconds
(without alignments)
16520.996 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_2501

Perfect score: 2351

Sequence: 1 ccttcaggaaaactcgtgtc.....attgaggagacgcgaagacga 2351

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_cm:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2351	100.0	3442	9	AB008430	Homo sapi
2	1351	57.5	3902	10	BC030329	Mus muscu
3	728.8	31.0	3719	10	BC009153	Mus muscu
4	719.6	30.6	3997	9	AB018336	Homo sapi
5	512.6	21.8	2605	9	BC021301	Homo sapi
6	478.8	20.4	2632	10	BC004009	Mus muscu
7	342.8	14.6	2888	9	HS080268	Homo sapi
8	312.8	13.3	2431	10	BC027077	Mus muscu
9	254	10.8	96183	9	AL161896	Human DNA
10	254	10.8	106578	9	AL137249	Human DNA
11	254	10.8	152053	2	AC022669	Homo sapi
12	229.6	9.8	2538	3	AY060484	Drosophila
13	204.2	8.7	224310	2	AC094777	Rattus no
14	199.4	8.5	221502	2	AC122885	Mus muscu
15	194	8.3	22595	9	AF156225	Homo sapi
16	193.4	8.2	152053	2	AC022669	Homo sapi
17	192.4	8.2	2867	9	HUMELIA	Human struc
18	192.4	8.2	3043	9	HUMEMP41	Homo sapien
19	189.4	8.1	5942	10	MUS41SP	Mus musculu
20	187	8.0	2882	10	AF106702	Mus muscu
21	185.2	7.9	6102	10	AB019256	Rattus no
22	183.6	7.8	5855	10	AB019257	Rattus no
23	181.8	7.7	3619	9	BC010674	Homo sapi
24	181.8	7.7	3643	9	HUMPTYPH	Human prote
25	181.2	7.7	2590	4	AF222767	Bos tauru
26	181.2	7.7	3661	10	AB032828	Rattus no
27	181.2	7.7	4543	10	AB032827	Rattus no
28	180.4	7.7	2625	9	AY049789	Homo sapi
29	180.4	7.7	3534	9	BC013885	Homo sapi
30	180.4	7.7	6263	9	AB002336	Human MRN
31	177.8	7.6	5921	3	DROP41A	Drosophila
32	177.2	7.5	3872	10	BC034751	Mus muscu
33	176.6	7.5	2964	10	AF044312	Mus muscu
34	174	7.4	3320	10	AF061283	Mus muscu
35	172.4	7.3	2758	5	XELCSK	X. laevis cy
36	169.6	7.2	3070	9	AY047584	Homo sapi
37	169.6	7.2	3925	9	BC034718	Homo sapi
38	169.6	7.2	4336	9	AF027299	Homo sapi
39	168.6	7.2	5097	5	AY124488	Danio rer
40	166.8	7.1	3451	10	AB032366	Mus muscu
41	165	7.0	4051	10	AF152247	Mus muscu
42	162.6	6.9	14917	6	AX375624	Sequence
43	162.4	6.9	393	11	G60059	SHGC-130885
44	161.8	6.9	2717	10	AF177146	Mus muscu
45	161.6	6.9	3309	9	AF069072	Homo sapi

ALIGNMENTS

RESULT 1	AB008430	AB008430	3442 bp	mRNA	linear	PRI 13-FEB-1999
LOCUS	Homo sapiens	Homo sapiens	mRNA for CDEP, complete cds.			
DEFINITION	AB008430	AB008430				
ACCESSION	AB008430.1	GI:2766164				
VERSION	CDEP.					
KEYWORDS	Homo sapiens	embryo cartilage chondrocyte cDNA to mRNA.				
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (sites)					
	Koyano, Y., Kawamoto, T., Shen, M., Yan, W., Noshiro, M., Fujii, K. and Kato, Y.					
TITLE	Molecular cloning and characterization of CDEP, a novel human					

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors
Biochem Biophys. Res. Commun. 241 (2), 369-375 (1997)
98086358
2 (bases 1 to 3442)
Koyano,Y., Kawamoto,T. and Kato,Y.
Direct Submission
Submitted (22-Oct-1997) Takeshi Kawamoto, Hiroshima University School of Dentistry, Department of Biochemistry; 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734, Japan
(E-mail:tkawano@ipc.hiroshima-u.ac.jp, Tel:082-257-5688, Fax:082-257-5629)

FEATURES

Location/Qualifiers
1..3442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="chondrocyte"
/tissue_type="cartilage"
/dev_stage="embryo"
49..3186
/function="Rho Guanine Nucleotide Exchange Factor"
/note="Band 4.1 superfamily"

CDS

/codon_start=1
/product="CDEP"
/protein_id="BAA24267.1"
/db_xref="GI:2766165"
/translation="MGELEORPTGSRILGAPENSISTLERCQKPPPTSPGKLVSIKI QMLDITQAEFVQPAKGVLLDVAENHLNVEGDI FGLEFPDHRKRLTWNLDLLAPIV KOIRRPKHVVFVKFFPDHDTLOEBLRYFALQVKQDLAQGLTNCNDTSAALLI SHVQSEIGDFEALDREHLAKNKYIQODALEDKIVFEFHNNHIGQTPAESDFOLLEI ARRLMEYGIIRLPKADREGTKINLAVANTGILVFOGFKINAFNNAKVPKLSFKRKF LKLRPDANSAYODTLEFLMASRDFCKSEFWKICVEHHAFFRLEFEPKPKPKPVPFSRG SSFSPSGTKOVLVDYVKEGGHKVQFERKHSKIHISRLASQPTSELSEVLEQSQQS TSELTGEGAESGSGSCRGKPKVSGEPGSPAPRSPGAKNQADGAASAPTEE EEEVVKDRTQSKPPQPPSTGSLTSGPSHLSVNSQGGVAPANTVSPNLSPTDK ASPLSPLENDQACPTDDEDEGRKRFTDKRAYFAKEVSTERTYI LKDEVIITSWF QSTVSKEDAMPALAKSLIFPNPEFLKHFNTFLKEIQRLALWEGRSNAQIRDYQRI G DVMKNIQCMKHLAHLNKHSEALEALENGIKSSRLENFCRDFLOKVCYUPLNTFL LRLHLRHVKOVLRLCKHPPSHADPRDCAALAEITEMVAQLHGTMIKWNFOKL HELKDLGIDNLVVGREFIRLGSLSKLSKGLQORFFLENDVLLLTSLRLTASNQ FKVGQLPIYMTIESEDEWGPCHLTILRGQRSLIIVAASRSSEMERKVEDIQMAID LAEKSSPAPELASPPDNKSPDEATAQDEDDLSASRTSLERQAPHRGNTMVHV CWHRTSYSMVDFSTAVENQLSNLLRKFKNSNGWQKLWVTFNCLFFYKSHQDNHP LASPLLYGSLTIPSESENIQDYVFKLHFKSHVYFRAESEYTFERWNEVIRSATS ASRPHVLSHKSLEYV"
polyA_site 3442
/note="50 a nucleotides"
BASE COUNT 864 a 952 c 927 g 699 t
ORIGIN

Query Match 100.0%; Score 2351; DB 9; Length 3442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTCTCAGAAAAC	TGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT	60
Dbl	151	CTTTCAGAAAAC	TGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT	210
QY	61	GAAGTTCCCAAG	AGCTCTCTGGGAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC	120
Dbl	211	GAAGTTCCCAAG	AGCTCTCTGGGAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC	270
QY	121	CTCGTGAAGGTG	ACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAAGATCACGGTGTGG	180
Dbl	271	CTCGTGAAGGTG	ACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAAGATCACGGTGTGG	330
QY	181	CTGGATCTCCTA	AAACCCATTGTTGAACAGATTAGAGGCCAACAGCAGCTGTTGTTAAAG	240
Dbl	331	CTGGATCTCCTA	AAACCCATTGTTGAACAGATTAGAGGCCAACAGCAGCTGTTGTTAAAG	390
QY	241	TTTGTGGTGAAT	TCTTTCGGCTTGACACACACACTCAAGAAAGAACTCACAAGGTAC	300
Dbl	1471	ACTGGCAGTCTC	ACCTTTCCGAGCTGTCTGTGAACTCGACGGGGGAGTGGCCCTTGCC	1530

QY	1381	AACGTGACCTTGCTCCCAACCTGAGCCCGACACCAAGCAGGCCTCTCCCTTTGATCAGC	1440
Db	1531	AACGTGACCTTGCTCCCAACCTGAGCCCGACACCAAGCAGGCCTCTCCCTTTGATCAGC	1590
QY	1441	CCGCTGCTGAATGACCAAGCCTGCCCGGACCGACGATGAGGCGCGGAGGAG	1500
Db	1591	CCGCTGCTGAATGACCAAGCCTGCCCGGACCGACGATGAGGCGCGGAGGAG	1650
QY	1501	AGATTCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTCTACACCGAGCGGAACA	1560
Db	1651	AGATTCCCAACTGATAAGCGTACTTCATAGCTAAGGAAGTCTACACCGAGCGGAACA	1710
QY	1561	TATCTGAAGATCTGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAGCAAGAGGAC	1620
Db	1711	TATCTGAAGATCTGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAGCAAGAGGAC	1770
QY	1621	GCCATGCCGAAGCACTGAAAGCTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT	1680
Db	1771	GCCATGCCGAAGCACTGAAAGCTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT	1830
QY	1681	CATACTAATTTTCTCAAGGAAATGAGCAACGACTTGCCCTGTGGGAAGCCGCTCAAAAT	1740
Db	1831	CATACTAATTTTCTCAAGGAAATGAGCAACGACTTGCCCTGTGGGAAGCCGCTCAAAAT	1890
QY	1741	GCCCAATCAGAGATTACCAAGAAATCGGCGATGTCTATGCTGAAGAACATTCAGGGCATG	1800
Db	1891	GCCCAATCAGAGATTACCAAGAAATCGGCGATGTCTATGCTGAAGAACATTCAGGGCATG	1950
QY	1801	AAGCACTGCGGCTCACCTGTGGAAGCACAGGAGCCCTTGAGGCCCTGGAGATGGA	1860
Db	1951	AAGCACTGCGGCTCACCTGTGGAAGCACAGGAGCCCTTGAGGCCCTGGAGATGGA	2010
QY	1861	ATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGAGACTTTGAGCTGCAGAAAGTGTGT	1920
Db	2011	ATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGAGACTTTGAGCTGCAGAAAGTGTGT	2070
QY	1921	TACCTACCGCTCAACACTTCTCTCGGGCCACTGACCGGCTCATGCACTACAGCAG	1980
Db	2071	TACCTACCGCTCAACACTTCTCTCGGGCCACTGACCGGCTCATGCACTACAGCAG	2130
QY	1981	GTCTTGAGGCGCTGTGCAAAACACCCGCGGACGACCCGCTTCAGGGACTGCCGA	2040
Db	2131	GTCTTGAGGCGCTGTGCAAAACACCCGCGGACGACCCGCTTCAGGGACTGCCGA	2190
QY	2041	GCGCTTTGGCAGAGATCAGGAGATGGTGGCAGCTCCACCGTACGATGATCAAGATG	2100
Db	2191	GCGCTTTGGCAGAGATCAGGAGATGGTGGCAGCTCCACCGTACGATGATCAAGATG	2250
QY	2101	GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTTGATTGGCATTTGAAATCTTGTG	2160
Db	2251	GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTTGATTGGCATTTGAAATCTTGTG	2310
QY	2161	GTTCGCGGAGGAGTTCATCCCTCTGGGCGAGCTCAGCAAGCTTCGCGGAGGGGCTC	2220
Db	2311	GTTCGCGGAGGAGTTCATCCCTCTGGGCGAGCTCAGCAAGCTTCGCGGAGGGGCTC	2370
QY	2221	CAGCAGCGATGTTCTCTCTGTTCAACGAGCTCCTGCTATACACGAGCGGGGCTGACG	2280
Db	2371	CAGCAGCGATGTTCTCTCTGTTCAACGAGCTCCTGCTATACACGAGCGGGGCTGACG	2430
QY	2281	GCTTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAG	2340
Db	2431	GCTTCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAG	2490
QY	2341	AGCGAAGACGA 2351	
Db	2491	AGCGAAGACGA 2501	

RESULT 2
BC030329
LOCUS

BC030329 3902 bp mRNA linear ROD 07-AUG-2002

BASE COUNT

982 a 1083 c 997 g 840 t

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

FEATURES
source

CDS

Mus musculus, clone IMAGE:5376197, mRNA, partial cds.
BC030329
GI:20987935

house mouse.

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
1 (bases 1 to 3902)

Direct Submission
Submitted (07-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Calithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 65 Row: m Column: 19

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

Location/Qualifiers

1..3902

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="FVB/N-3"

/clone="IMAGE:5376197"

/tissue_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month

old mouse. Taken by biopsy."

/clone_lib="NCI_CGAP_Mam2"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1..2510

/codon_start=3

/product="Unknown (protein for IMAGE:5376197)"

/protein_id="AAH30329.1"

/db_xref="GI:20987936"

/translation="TRPEVARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTK
INAFNWKVKLSFKRFLIKLRDPDVASSYQDTLEFLMAGDFCKSPFKICVEHAP
FLFEFPKPKPVLFRSGSPFRGRTQKQVLDIVKGGHKVKQFERKHSKHSTRS
LVSPQAPNSPVPKSPQSPASLTFGETSPGGQSQQAKETKACTLELGHQSPALP
KSPGSKAADGTVPVPEEEEGKDGIRPSNPQPPQSTGSLTSPHLSELINS
QGGAAANVTSPNLSPDNKQASPLISPLNDQACPRTDDEEGRKRFPPTDKAYIA
KVSVTERTYLDLEVIASFQSTVSKDSMPKSLTFPFPFLHFTFTFLAKEIE
ORLALWEGRSNAHVYGDVQRIDVLMKNIQGMKHLAHLWKHSEALELTSIKSRR
LEHFCRDFELQKVYLPNTFLRPLRHLRHHYKHLERLCKHPNHDPRCDRAALA
RITENVAOLHGTMTIKENFQKHLKDLIGDNLVTPGREFIRLGSLSKLGLQOQ
EMFFLNDVLLYTSRGLTASNQFKVHGQPLPYMTIESEEWGPHCITLRGQRSI
IVAASSRSEKMEWMEDIQMIDLAESKNGPTPELLASSPDNKNKSPDEATADESEDD
LSASRTSLERQAPRGNTMVHVHRSTSVSMVDFSIAVNOLSGNLLRKFKNSNGW
KLWVYTFNFCFFYKSHQSDHPLSLPLGLYSITPSESENIHKDYVFLHFKHSHVYY
FRAESEYTFERWNEVIRSATSSASRAHILSHKESHLY"

ORIGIN

Query Match 57.5%; Score 1351; DB 10; Length 3902;
Best Local Similarity 84.7%; Pred. No. 0;
Matches 1542; Conservative 0; Mismatches 270; Indels 9; Gaps 2;

QY 540 GTCTCTAGAGATTGCGCGCTGAGAGATGATGGAATCCGGTTGCACCCGCGCAAGGA 599
DB 5 GCGTCGGAGGTGGCCGCTGCGCTAGAGATGACGGAATCCGGTTACACCCCTGCCAAGGA 54

QY 600 CAGGGAAGCAGCAGATCAATCTGGCGCTTGCACACGCGGAATCTAGTGTTCAGG 659
DB 65 CAGGGAAGCAGCAGATCAATCTGGCGCTTGCACACCTGGAATCTAGTGTTCAGG 124

QY 660 TTTCACATAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTTGAGCTTCAAGAGGA 719
DB 125 CTTTCAGGAAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTTGAGCTTCAAGCGGA 184

QY 720 GCGCTTTCATCAAGCTCCGGCCAGATGCCAATAGTGTGCTACCAAGATACCTTGAAT 779
DB 185 GCGCTTTCATCAAGCTCAGACCTGACGTGAATAGCTCATACCAAGACACCTTGGAGT 244

QY 780 CTTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAATACTGTGTGAACATCATGC 839
DB 245 CTTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAATACTGTGTGAACATCATGC 304

QY 840 CTTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCGCTCTTTAGCCGGG 899
DB 305 CTTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCGCTCTTTAGCCGGG 364

QY 900 GTCTCATTTTCAGTTCAGTGGCTCAGAACGAGGTTCTCGACTATGTTAAAGAAGG 959
DB 365 GTCTCATTTTCAGTTCAGTGGCTCAGAACGAGGTTCTCGACTATGTTAAAGAAGG 424

QY 960 AGCATAAGAAGTGCATGTTGAAAGAGCAGACAGATTCATTCATCCGAGCCT 1019
DB 425 CGGCCACAAGAAGTGCATGTTGAAAGAGCAGACAGATTCATTCATCCGAGCCT 484

QY 1020 TGCCTTCACAGCTTACAGACTGAATTCGGAAGTGTGGAGAGCTCTCAGCAGACCAAG 1079
DB 485 GTCTTCACAGCTTACAGACTGAATTCGGAAGTGTGGAGAGCTCTCAGCAGACCAAG 544

QY 1080 CTTTACATTTGGAGAGTGCAGATCTCCAGGGGCCAGAGTCCGCGCAGGAAAGGA 1139
DB 545 CTTTACATTTGGAGAGTGCAGATCTCCAGGGTGSCCAGAGCTGCCACCAAGCAAGGA 604

QY 1140 ACCGAGGTTTCGCGCGGAGCGGGTGCACCGAGCCCTGCGCCAGAGAGAGCC 1199
DB 605 AACAAAAGCTGTACTTGGAGTTGGGACCTCACCAAGCCCGCCCTCCCTAAGAGCC 664

QY 1200 CGCGGTAAACAGCAGCGGAGCGGAGCCGCTCGCGGCCACGAGGA- ---GAGGA 1253
DB 665 TCAGGGACCAAGCTGCAGATGGAAACACCGTGTGGTGGCCCTGAGGAAGAAGAGGAGGA 724

QY 1254 GGAGTGTGTTAAGGATPAGGACCCAGCAGAGTAAACCTCAGCCCGCCAGCCCAAGCAGG 1313
DB 725 AGAGGTGCGCAAGGAGCGCATCCGCGCAAGTAACTCAGCCCGCCAGCCCAAGCAGG 784

QY 1314 CTCCTCTGACTGGAGTCTCAGCTTTCCAGCTGTCTGTGAATCTGCAGGGGGAGTGGC 1373
DB 785 CTCCTCTGACTGGAGTCTCAGCTTTCCAGCTGTCTGTGAATCTGCAGGGGGAGTGGC 844

QY 1374 CCTCTGCAAGCTGACCTTCTCTCCCAAGCTGAGCCCGCCAGCAGCAGCCCTCTCCCT 1433
DB 845 CCAGCCCAAGCTGACCTTCTCTCCCAAGCTGAGCTGTGACACCAAGCAGCCCTCTCCCT 904

QY 1434 GATCAGCCCGCTGCTGAATGACCAAGCCCTGCCCGGAGCGAGTGAAGTGAAGCGCG 1493
DB 905 GATCAGCCCGCTGCTGAATGACCAAGCCCTGCCCGGAGCGAGTGAAGTGAAGCGCG 964

QY 1494 GAGGAAGATTTCCCAAGCTGAAGCGTACTTTCATAGTGAAGAGTGTCTACCAAGGA 1553
DB 965 GCGAAAGATTTCCCAAGCTGAAGCGTACTTTCATAGTGAAGAGTGTCTACCAAGGA 1024

QY 1554 GCGAACATATCTGAAGGATCTCGAAGTTATCACTTCGTGTTTCAGAGCACAGTGAAGCA 1613
DB 1025 GCGCAGTACCTGGAAGGACCTTGAAGTCAATTCGTGTTTCAGAGCACAGTGAAGCA 1084

QY 1614 AGAGGACGCAATGCGCGAAGCACTGAAAAGTCTCATATATCCCGAAATTTGAACCTTTGCA 1673
DB 1085 AGAGGACCTCCATGCGCGAAGCCTTGAAGAGTCTCATATTTCCCAATTTGAACCTTTGCA 1144

QY 1674 CAAATTTCTACTAATTTTCTCAAGGAAATGAGCAACGACTTGCCTGTGGGAAGCGG 1733
DB 1145 CAAATTTCTACTAATTTTCTCAAGGAAATGAGCAACGACTTGCCTGTGGGAAGCGG 1204

QY 1734 CTTCAATGCCCAATCA ---GAGATTACCAAGAAATGCGGATGTCATCTGAAGAACAT 1790
DB 1205 CTTCAATGCCCAATCAAGGAGATTTACCAAGAAATGCGGAGCTCATCTGTAAGAACAT 1264

QY 1791 TCAGGCGATGAAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGAGGCTTGGAGGCCCT 1850
DB 1265 TCAGGCGATGAAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGAGGCTTGGAGGCCCT 1324

QY 1851 GGAGAAATGAATCAAGAGTCCCGGGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCA 1910
DB 1325 GGAGAACTCCATCAAGGCTCGCGGCTGGAACATTTCTGCGAGACTTGCAGCTGCA 1384

QY 1911 GAAGGTGTCTTACCTACCGCTCAACACCTTCTCTGCGGCGCACTGCCCGGCTCATGCA 1970
DB 1385 GAAGGTGTCTTACCTACCGCTCAACACCTTCTCTGCGGCGCACTGCCCGGCTCATGCA 1444

QY 1971 CTACAAGCAGGTCCTGGAGGCTGTGCAAAACACACCGCGAGCCAGCGGACTTCAG 2030
DB 1445 CTATAAGCATGCTCTGGAGAGGCTGTGCAAGCACACCCACCAACCGCGGACTTCAG 1504

QY 2031 GGAGTCCGAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAGCTCCAGGTTACAT 2090
DB 1505 GGAGTCCGAGCGGCTTTGGCAGAGATCACGAGATGTTGGCAGAGCTCCAGGTTACAT 1564

QY 2091 GATCAAGATGGAAGATTTCCAGAACTGCACAACTCAAGAAAGATTTTCATTTGGCATCA 2150
DB 1565 GATCAAGATGGAAGATTTCCAGAACTGCACAACTCAAGAAAGATTTTCATTTGGCATCA 1624

QY 2151 CAATCTTGTGTTCCGGGAAGGAGTTTCATCGTCTGGCAGCGCTCAGCAAGCTCTCGG 2210
DB 1625 CAATCTTGTGTTCCGGGAAGGAGTTTCATCGTCTGGCAGCGCTCAGCAAGCTCTCGG 1684

QY 2211 GAAGGGCTTCAGAGCGGCTGTTCTTCTGTTCACAGAGTCTCTGCTATACAGAGCG 2270
DB 1685 GAAGGGCTTCAGAGCGGCTGTTCTTCTGTTCACAGATGTTCTGCTGTATACAGCG 1744

QY 2271 GGGGCTGAGCGCTCCAATCAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGAT 2330
DB 1745 GGGGCTGAGCGCTCCAATCAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGAT 1804

QY 2331 GATTGAGGAGCGGAAGCA 2351
DB 1805 GATCGAGGAGGTGAGGAGGA 1825

RESULT 3
LOCUS BC009153
DEFINITION Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304
IMAGE:2655209, mRNA, complete cds.
ACCESSION BC009153
VERSION BC009153.1 GI:14318718
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3719)
AUTHORS Strausberg,R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 5 Row: m Column: 20.

FEATURES

Location/Qualifiers

- 1..3719
- /organism="Mus musculus"
- /db_xref="taxon:10090"
- /map="FVB/N-3"
- /clone="MGC:6304 IMAGE:2655209"
- /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."
- /clone_lib="NCI_CGAP_Mam2"
- /lab_host="DH10B"
- /note="Vector: pCMV-SPORT6"
- 99..3296
- /codon_start=1
- /product="similar to KIAA0793 gene product"
- /protein_id="AAH09153.1"
- /db_xref="gi:14318719"
- /translation="MGTEGYRALPTSGRLGGTATGVSTLPEQSLSPRMOEKHM RIRKVLDPSTNAFLDIPEKQDQVLLTOVWHLNLECDYGLFKNVOSYWIWLEPM KPIIROVBKNEVLRLAVKFPDPDQLOEYTRYLEALQKRLDRLERLITCTANTA ALLSHLQSEIGDEYDLREHLKANLEYLPNOESLEKILDFHQRHTGQTPAESDFQ VLRTARKUDELGRFHMADREGTKINLAVSHMGVLVFEQTKTINFNWSKRLSFK KRELLKHPVHPQIDTLEFLILGSDCKENFWKICVEYHTFRLSDQPKAKAVF FSRGSSFRSRTQKLVYVKGDMKRPYERRHSKRTSLHALTVLDPKOSVSFTD GLRTASLSANLVFPYPPSSSLPPLNLDKSSSLVDPAVPKISTAAERSGSPS SDDGSTOSALHCPVPLRPGFSDPSQPSLSKSHLSLCPQLAALTAEOGAS PVLSPVLSGATARNQOEKHKHMPDEAFYFAKEILATERYLYKDLVITYWFRS VLIKEEAMPALMALLFNSNDIPVYEFHFRGLHEVEQRLALWEGPSSAHLKGDHQRIG ILLRNMQLKFTSYFQRDEVLTELEKATKCKKLEAVYKEFELOKVCYLPNLFL KPVORLHYRLLSRLCAHYSFGHRDYADCHEALKAITVETTELQOQSLRLENLQKLT ELORDLVGNLAPGREIFREGCLHKLTKKLOQRMFFLFSDMLLYTSKSVTGASHF RIRGLPLGRMLVESENEWSVLHCFTIYAQKTIYVAASRLKEKWMQDLNRAIQ A KTTGDSPPVLLGPVYTRPRSDSEVLEESDGRNGRSLSEGNQRHANTMHVC YRNTSVSRADHASVENOLSYLLRKFNSNGWKLWVFTNFCFLFFYKTHODDYPLA SLPLIGVSVPLEADSIHKDYVFKLPKSHVYFFRAESKYTFERWMDVIKRASSPG RPPSTQCSHHSGLEAIREKEACPSCLDKNL"

CDS

997 a 979 c 929 g 814 t

BASE COUNT
ORIGIN

Query Match 31.0%; Score 728.8; DB 10; Length 3719;
Best Local Similarity 58.5%; Pred. No. 4.8e-165;
Matches 1374; Conservative 0; Mismatches 947; Indels 27; Gaps 5;

QY 10 AAACGTCTCCATCAAAATCCAGATGCTGGATGACACCCAGGACATTTGAAGTTCA 69
Db 222 AAGCACATCGCATGAGAGTGTGGACAGCATCTGTAGATTAATTCACATCGAG 281
QY 70 CAAAGAGCTCTCGGAAGTGTCTGCTGGATGAGTTCGAAACCACTCAACCTCGTGGAA 129
Db 282 CCGAAGTGTGACGGCAGGTCTTACTGACTCAAGTGTGGAAGCANTTTGAATCTCATGAA 341
QY 130 GGTGACTATTTTGGCTCGAGTTTCTCTGATCACAAAAAGATCACGGTGTGGTGGATCTC 189

Db 342 TGTGACTACTTTGGTCTGGAGTTCAAGAATGTCCAGTCTCTACTGGATTGGTGTGAACCT 401
QY 190 CTAATAACCCATTGTGAACACAGATTAGAAGGCCAAAGCACGTTGTTTAAAGTTTGTGGTG 249
Db 402 ATGAATAACCCATCATATTAGGCAAGTACGAAAGCCAAATAATTCGCGTGTCTCGCTGGCAGTA 461
QY 250 AAATTTCTTTCCGCTGACACACACAACTCCAAAGAGAACTCACAAAGTACCTGTTTCGGG 309
Db 462 AAATTTCTTTCCGCTGACACACAACTCCGTCAGTGTCAAGAGAGTACACAAAGTACCTGTTTCGG 521
QY 310 CTGCGAGTGAAGCAGCAGGTGGCTCAAGCCAGGTGACGTGTAAAGTACACACAGCAGCT 369
Db 522 TTGCAACTCAAGAGAGACCTCTCGAGGAAACGCTGTGACGCGCCACACTTCGAGCC 581
QY 370 CTTCTGATTTACACATCTGCAATCTGAGATTGGGATTTTGGATGTAAGCCCTTGGACAGA 429
Db 582 CTTCTCATATCCACCTCTTCAGTCGGAATTCGGAGATTATGATGAAACCTGGATCGA 641
QY 430 GAGCACTTAGCAAAAAATAATACATACCTCAGCAAGACGCACTAGAGGACAAAAATCGTG 489
Db 642 GAACACCTCAAGGCCAATGAATACTTGCCCAACCCAGGAGAAATCTCTAGAAAAGATACTA 701
QY 490 GAATTTCCACATAACACATATTGGACAAACACACAGCAGATCAGATTTCAGCTCTTAGAG 549
Db 702 GACTTCCATCAGAGGCACACGGGCCAGACTCCTGCAGAGTACAGATTTCAGAGGTGCTTGAG 761
QY 550 ATTGCCGCTGCGCTAGAGATGTATGGAATCGGTGTGCACCGGCCCAAGCACAGGGAAGCC 609
Db 762 ATTGCAAGAACTGGAATGTATGGCATCAGGTTCACATGGCTTCTCAGAGAGAAAGG 821
QY 610 ACGAAGATCAATCTGGCCGTTGCCAACACGGAATTTCTAGTGTTCAGGGTTTCTACTAAG 669
Db 822 ACCAAGATTAACTAGCAGTTTCTCACATGGGTGCTCTGTTCCAGGTTACCACCAA 881
QY 670 ATCAATGCTTCAACTGGGCGTCCAGGTGCGGAAGCTGTGCTTCAAGAGGAAGCGCTTTCTC 729
Db 882 ATCAATACCTTCAACTGGTCCAAAGTCCGTAACCTAAAGCTTCAAGAGAAAAAGTTCTT 941
QY 730 ATCAAGCTCCGCGCAGATCCCAATAGTGGTACCAAGATACCTTGAATTCCTGTAGGCC 789
Db 942 ATCAAACTCCACCCCTGAGGTTCATGGGCCCTACCAAGACACGTTAGTGTGTTGGGT 1001
QY 790 AGTCGGGATTTCTGCAAGTCTTCTGGAAAACTCTGTGTTGAACATCATGCTCTTTTGA 849
Db 1002 AGCAGAGATGAATGTAAGAATCTCTGGAAGATATGTGTGAGTACCATACCTTTTGA 1061
QY 850 CTTTTTGAAGCCCAAAACCAAGCCCAAGCCGCTCTTTAGCGGGGGTCAATCATTT 909
Db 1062 CTCTCTGACGCTTAAGCCAAAGGCAAGGCTGTCTTCTCAGCGGAGGCTCTCTCTTC 1121
QY 910 CGGTTTCAGTGGTGGACTCAGAGCAGGTTCCTCGACTATGTTAAAGAGGAGGACATNAG 969
Db 1122 AGATACATGGAAGAACTCAGAAACAACTAGTAGATTATGTCAAGACGGTGAATGAAG 1181
QY 970 AAGGTGCACTTTGAAAGGAAGCACACAGCAAGATTC---ATTCTATCCGAGCCTTGCTTCA 1026
Db 1182 AGAATTCATACGAAGACGGCACAGTACAGTACGACATCTCTTCATGCTCTGACTGTA 1241
QY 1027 CAGCCTTACAGAACTGAATTTGGAAAGTGTCTGGAGCAGTCTTCAGCAGAGCACCAGCCTTACA 1086
Db 1242 GATCTGCTTAAACAGAGCGCTCTCTTACCCGATGGCTTTCAGAACTTCTGCTCCCTGCTCT 1301
QY 1087 TTTGGAGAGGTGGCCAAATCTCCAGGGGCCAGAGCTGCGCGGAGGAGAAAGAACCGAAG 1146
Db 1302 TCAGCAAAATGTCTCTTTATCC-----ACCCCTAGTCTTCTCTTTGTGCTCTCC 1351
QY 1147 GTTTCCGCGGGGCGCGGGTCCGACCCGAGCCTCTGCGCCGAGGAGAAAGCCCGGGT 1206
Db 1352 TGGTCTGCCAATTTCAAGGACAGCAGCAGCTCCCTTGTGGATCCCGAGGCTCCCGTCTAT 1411
QY 1207 AACAAGCAGCGGACGGAGCCGCTCGGCCCCACCGAGGAAGAGGAGGTGCTTTAAG 1266

Db 1412 CAAGACACAGCAGCAGAGAGGAGGAGCGGAGCCATCATC--ATCTGAGCGGCCACGACACA 1469
QY 1267 GATAGACCCACGAGATAAACCTCAGCCCCGCGAGCCAAAGCACAGAGCTCCCTGACTGGC 1326
Db 1470 CAGTCGGCCCATCTCCCTGGACCCCGGTGCTCCCGGCTGGTCCAGGCTTTTCTATGAT 1529
QY 1327 AGTCCTCAGCTTTCCGAGCTGTCTGTAACTCGCAGGGGGAGTGGCCCTGCCAACGTG 1386
Db 1530 AGTCCTCAGCTTTCCCTCCAGCTGAAGAGCCACCTGAGCTGTGCCCTGAGCTTCAG 1589
QY 1387 ACCTTGCTCCCAACCTGAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGCCCGCTG 1446
Db 1590 GCCGG-----CTGAGTACAGCTGAGCAGGGTGATCCCTGTGCTCAGCCCTGTG 1640
QY 1447 CTGAATGACCGGCTGCCCGCGAGGAGCATGAGGATGAGGCGCGGAGGAAGATTC 1506
Db 1641 CTCAGTGGCGCTGTACAGCCAGGATGACCAACCAAGAAAGACAGCAACACAGCACATG 1700
QY 1507 CCAACTGATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACCCAGCGAGCGACATATCTG 1566
Db 1701 CCAGAAGATGAGCCCTATTTCATACCCAAGGAGATTCTCGCTACAGACGACCACTATCTG 1760
QY 1567 AAGGATCTCGAAGTATCATTCTGCTGGTGTTCAGACACAGTGCAGAAAGAGACGCCATG 1626
Db 1761 AAGGATTTAGAAGTTATCATTCTGTGTGTTCCGGAGCGTGTGATCAAGGAGGAGGCCATG 1820
QY 1627 CCGGAGCAGCTGAAAGTCTCATATTCCGGAATTTGAACCTTTGCACAAATTTTCATCT 1686
Db 1821 CTTGAGCGCCCTGATGGCCCTGCTTTTCTCCAATTTGATTCAGTCTACGAGTTCACACAGA 1880
QY 1687 AATTTTCTCAAGAAATGAGCAAGCACTTGCCTTGTGGGAGGCGGTCAAAATGCCCA- 1745
Db 1881 GCCTTCTCACGAGTGGAAACAGAGGCTGGCACTCTGGGAAGGCGCTCCAGTCCCCAC 1940
QY 1746 --AATCAGAGATTACAAAGAAATCGCGATGTCTAGTGTGAAGAAATTCAGGGCATGAAG 1803
Db 1941 TTAAGAGGTGATCACCAGCAGATCGGGACATCTCTCTCAGGAACATGCGTCAGTTAAG 2000
QY 1804 CACCTGGGCGCTCACTGTGGAAGCACAGCAGGCGCTTGGAGGCGCTCGAGAATGGAATC 1863
Db 2001 GAATTTACTAGTACTTCCAAGACACAGATGAGTGCTTAACAGAACTGGAAAGGCCACA 2060
QY 1864 AAGAGCTCCCGCGCTGGAGAACTTTCGACGAGACTTTAGCTGCAGAAAGTGTGTAC 1923
Db 2061 AAACACTGTAAAGAGCTGGAGCAGCTTACAAGAGTTTGAAGTCCAAAAGTCTGCTAC 2120
QY 1924 CTACCGCTCAACACTTCTCTCTGGGCGCAGTGCACCGGCTCATGCATCAGCAGGTC 1983
Db 2121 CTGCTCTCAACACATCTCTGCTGAAGCCCGTCCAGAGGCTAGTCCACTACCGTCTGCTG 2180
QY 1984 CTGGAGCGGCTCTGCAAAACACCACCGCGGAGCCAGCGGCTTACAGGAGCTGCCGAGCC 2043
Db 2181 CTGAGCGGCTGTGTCTACTACTTCTCTGGGACCGGAGTATGCCGACTGCCCATGAG 2240
QY 2044 GCTTTGGCAGAGATCACGGAGATGTTGGCACAGCTCCACGCTACGATGATCAAGATGGAG 2103
Db 2241 GCACCTGAAGGCCATCACAGAAGTGAACCCAGCTCCAGCAAGGCTTACC CGGCTGGA 2300
QY 2104 AATTTCCAGAAGCTGCAGAACTCAAGAAAGATTGATGGCATTTGACAACTTTGGT 2163
Db 2301 AACCTACAGAAATACCGSAGCTACAGCGAGACCTTGGTGTGAGAAACCTCATTTGCT 2360
QY 2164 CCGGAAGGAGTTCATCTGCTGGGACGCTCAGCAAGCTCTCGGGAGGGGCTCCAG 2223
Db 2361 CTGGAGGAGATTTATCCGTGAGGCTTCCCTGCAACAGCTCACCAGAAAGGGCCCTGCA 2420
QY 2224 CAGCGCATGTTCTCTCTTCAACAGCTCTCTGCTATACACGAGCGGGGCTGACGGCC 2283
Db 2421 CAGAGGATGTTTTTCTGTCTCAGATATGTTTACTGTATACAAAGCAAAAGTGTACAGGA 2480
QY 2284 TCCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATCAGCATTCAGGAGAGC 2343
Db 2481 GCCAGTATTTCCGGATCCGTGCTTCCCTTCACTCCGTGGCATCTGTTAGAAAGAAAGT 2540

QY 2344 GAAGACGA 2351
Db 2541 GAGATGA 2548
RESULT 4
AB018336
LOCUS AB018336 3997 bp mRNA linear PRI 16-JUN-1999
DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.
ACCESSION AB018336
VERSION AB018336.1 GI:3882306
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hk05692.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara.O.
Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
DNA Res. 5 (5), 277-286 (1998)
99087487
2 (bases 1 to 3997)
Ohara.O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
Direct Submission
Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)
FEATURES
Location/Qualifiers
1..3997
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hk05692"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
1..3997
/gene="KIAA0793"
118..3282
/gene="KIAA0793"
/codon_start=1
/product="KIAA0793 protein"
/protein_id="BAA34513.1"
/db_xref="GI:3882307"
/translation="MGIEGTyrVLOTAGMRLGAOTPVGVSTLEPGOTLLPRMOEKHL
HLRVKLLDNTMEIFDIEPKCDGOVLLTQVWKRLNLVECDYFGMEFQNTQSWINLEPM
KPIIRIRRNANVRLAVKFFPPDQQLQOEYTRYLFALQKRDLLERLUTCADTTA
ALSTSHLQSEIGDYDETLDREHLKNEYLPQQHCKLEKLEFHOKHVGGTQPAESDFQ
VLETKLEMYGIRFHMASDREGTKIQLAVSHMGVLVQGTGTTKINTFNWSVKRLSFK
RKRELIKLHPVHGPDYDTLEFLGSDCKENWKICVYHTFFRLDPOPKAKAVF
FSRGSFYSRCKTQKLVDFKDSMKRIPYERRSKHTSVRALTDLPKQTSFPE
GLRTPSPANNAFYSLSPSTLVPSGLPEPKDSSSLTDQOVSYVYKSPAAARRSGAVA
GGDTPSAQPLGFPALQPGELSTKSPQSPSSRSKSLDPAQVPLGPAEQSSPL
LSPVSDAGGAGMDCPEPRHKRVPAEAYFTVKEILATERYTKDLEVITVWFRSAV
KEDAMPATLMTLLFSDNIDPIYEFHRGFLREVEQRLALWEGFSKAKTSGHQRIQDILL
RNMQLKEFTSYFORHDEVLTELEKATCKKCKLEAVYKEFELQKVCYLPNTFLKPI
ORLLHYRLLLRLRGCHVSPGHHDYADCDKALKATTEVTTTLQHLIRLENLQKTELQ
RDLVGIENLTAPGREFIRESCLLHLTKKGLQORMFFLESDMLLTSKGVAGTSFPIR
GLPLQGLMVEESDNWSVPHCFITYAAQKTIIVVAASTRLKEKWKMDLDSAIQAAS
GGDTPALPGRTVCTRPSPNEVSEDESDDARGVRSLEGGHQRANTMTVCWY
RNTSVSRADHSAAVENOLSGVLLRKFKNHGWKLWVFTNCFLLFKTHODDYPAS
LPLGLSVSIPREADGIHKDYVFKLFQKSHVYFPAESKVTFRWMEVIQGAASSAGR
APSTVQDGPQPPSSGLEGMGRKEP"
BASE COUNT 998 a 1132 c 1068 g 799 t
Query Match 30.6%; Score 719.6; DB 9; Length 3997;

Best Local Similarity 58.3%; Pred. No. 8.1e-163; Matches 1372; Conservative 0; Mismatches 944; Indels 36; Gaps 5;	
Qy	6 AGGAAACCTCGTGTCCATCAAAATCCAGATCGTGATCACCCAGGAGCATTTTGAAGT 65
Db	237 AGAAGACACCTGCACCTCAGAGTAAGCTGCTGGACAACACATGGAATATTTGACAT 296
Qy	66 TCCAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCACTTTGCAACACCTCAACCTCGT 125
Db	297 TGAGCCTAAATGCGATGGCCAGGTATTACTGACACAAGTGTGGAAGCGTTTAAACCTGTT 356
Qy	126 GGAAGTGACATATTTGGCCTCGAGTTTCCCTGATCACAAGAGATCAGGTGTGGCTGGA 185
Db	357 AGAATGTGACTACTTCGGGATGGAGTTTCAAAATACTCAGTCTACTGATTTGGCTTGA 416
Qy	186 TCTCTAAACCCCATTTGTGAACAGATAGGAAGGCCAAAGCAGTGTGTGTTAAGTTGT 245
Db	417 ACCTATGAACCCCATATTAGGCAATACGAAGGCCAAGAATGTGGTGTCTCGCCTAGC 476
Qy	246 GGTGAATTTCTTCGCGCTGACCACACAACTCCAAGAAGAACTCACAAGGTACCTGTT 305
Db	477 TGTAAATTTTTCACCTGTATCTCGTGTGACGTACAAGAAGATATACAAGATACTTGT 536
Qy	306 CGCGCTGAGGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACACAGGC 365
Db	537 TGCTTTGCAACTTAAGAGAGACCTGCTGGAAAGCGTTTGACCTGTGCTGACACACAGC 596
Qy	366 AGCTCTCTTGATTTACACACATTTGCAATCTGAGATTTGGGATTTTGATGAAGCTTGA 425
Db	597 GGCGTCTCAGTCCCATCTCTGCACTCGGAATAGGAGATACGATGAAGACGCTGGA 656
Qy	426 CAGAGCAGCTTAGCAAAAAATAATACATACCTCAGCAAGCAGCACTAGAGGACAAAAT 485
Db	657 CCGAGAGCACCTCAAAGTGAAGAGTATTTCCCTGGCCAGCAGCACTGCCCTTGAGAGAT 716
Qy	486 CGTGAATTTTACCATTAACCATTTGGACAACACACAGCAGAACTCAGATTTCCAGCTCCT 545
Db	717 ACTAGAATTCCATCAGAAGCAGCTGGGCCAGACACCTGCTGAGTCGGATTTCCAGGTGCT 776
Qy	546 AGAGATTCCCGCTCGGTAGAGATGTATGGAATCCGGTTGCACCGCCAGGACACAGGA 605
Db	777 CGAATTGCTCGAAGTTGGAATGTACGGCATCAGATTTCAATAGGCTCTGACAGGGA 836
Qy	606 AGGCAGAGATCAATCTGGCCGTTGGCCAAACACGGGAATTCAGTGTGTTACGGTTCAC 665
Db	837 AGGAACCAAGATTCAACTGGCAGTTTCCCACATGGTGTACTCGTGTTCAGGCGCAC 896
Qy	666 TAAGATCAATGCCCTTCAACTGGCGCAAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGGTT 725
Db	897 CAAATCAACACTTTCAACTGGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
Qy	726 TCTCATCAAGCTCCGGCCAGATGCCAATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 785
Db	957 TCTTATCAAACTTCATCCAGAGGTTTATGAGCCTTACCAGGACACATATGAGATTTTGT 1016
Qy	786 GGCCAGTCGGGATTTCTGCAAGTCCTTCTGGAAATCTGTGTTGAACATCATGCTTCTT 845
Db	1017 GGGTAGTAGAGATGAATCTGAAGACTTCTGGAAGATTTGCTGGAGTATCACACCTTTT 1076
Qy	846 TAGACTTTTGAAGGCCCAACCAAGCCCAAGCCGCTCTCTTTAGCCGGGGGTATC 905
Db	1077 TAGACTTTTGACCAACCTTAAGCCAAAGCAAAAAGCCGCTCTTCTTCAGCCGGGCTCTC 1136
Qy	906 ATTTTCGGTTTCAAGTGTGCGACTCAGAAGCAGGTTCTCGACTATGTTTAAAGAGGAGACA 965
Db	1137 CTTCAATACAGTGGAGAAGTCAAGAACAACTAGTAGATTTATTTCAAAGACAGTGAAT 1196
Qy	966 TAAGAAGGTGCGATTTTGAAGGAAGACACAGCAAGATTCATTCATCGGAGCCTTGCTTC 1025
Db	1197 GAAGAGAATTCATATGAAGAAGGCACAGCAAGCCACACGTCCTTCTTCAGCCGGGCTCTC 1242
Qy	1026 ACAGCCTTACAGACTGAATTCGGAAGTGTGCTGGAGCAGTCTCAGCAGAGACACCGCTTAC 1085

QY 786 GGCCAGTCGGGATTTCTGGAAGTCTTCTTCTGAAATACTGTGTGAACATCATGCTCTTT 845
Db 1025 GGGTAGTAGAGATGAATAGAACTTCTGGAAGTGTGTGTGAGTATCACACCTTTT 1084
QY 846 TAGACTTTTGAAGAGCCCAAGCCCAAGCCGCTCTTTAGCCGGGGTCATC 905
Db 1085 TAGACTTTTGAAGAGCCCAAGCCCAAGCCGCTCTTTAGCCGGGGTCCTC 1144
QY 906 ATTTGGTTCAGTGGTTCAGAGCAGAGTTCGACTATGTTAAAGAGGAGACA 965
Db 1145 CTTTCAGATACAGTGGAGAACTCAGAACTAGTAGATATTCAAAGACAGTGAAT 1204
QY 966 TAGAAGGTTCAGTTGAAGAGGAGCAGCAGAGTATTCATTCGCGAGCCCTTGC 1025
Db 1205 GAAGAGAAATTCATATGAAGAAGGAGCAGCAGAGCCACACCTCGCTTCG----- 1255
QY 1026 ACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGCTCTCAGCAGACGACCTTAC 1085
Db 1256 -----AGCTGTGACTCGACCTTACCAGACAGAGATCTCATTTCCCGAGGAT 1306
QY 1086 ATTTGGAGAGGTGCGGAATCTCCAGGGGCCAGAGCTGCCGCGAGGAAAGAACCGAA 1145
Db 1307 GAGGACTCTCTCCCTCCCTCAGCGAATGCTTTTACTCGCTCTCTCTCCCTCACTCT 1366
QY 1146 GGTTCGCGCGGAGCGGGGTGCGACCCGAGCCCTGCGCCGAGGAGAACCCCGGG 1205
Db 1367 GGTCCCTCTGGCTGCGAGTTTAAGGACAGCAGCAGCTCCCTCACAGATCCCGAGT 1426
QY 1206 TAACAGCAGCGGAGCGGCGCTCGGCGCCAGGAGAGAGGAGGAGG-----TCGT 1262
Db 1427 TTCTAGCTCAAGTCCAGCTCGAGAGGCGCAGTGGAGCAGTGGCTGGAGGCCCGGA 1486
QY 1263 TAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCGCCGAGCAGCAGAGCCTCCCTGAC 1322
Db 1487 CACACCATCGGCCAGCCCTCGGCGCCCGCAGCTCCAGCCTGGTCCAGGCTTTCCAC 1546
QY 1323 TGGCAGTCTCAGCTTTCAGAGTGTCTGTGAAGTGTGCGAGGGGGGAGTGCCCTGCA 1382
Db 1547 GAAGAGTCTCTCAGCTTCTCCCTCCAGCGGAGAGCCCTGAGTCTGAGCCTGCATT 1606
QY 1383 CGTGACCTTCTCCCAACTGAGCCCGGACACCAAGCAGGCTCTCCCTTGATGACGCC 1442
Db 1607 TCAGTGTCTT-----TGGGCCCCAGCTGAACAGGGCTCATCCCTGAGGCC 1657
QY 1443 GCTGCTGAATGACAGGCTTCCCGGAGCGGAGCAGTGAAGTGAAGGAGGAGAG 1502
Db 1658 TGTCTCAGTGTGCTGGGAGCGCGGATGAGTGGAG-----GAGCCACAGACAGCG 1714
QY 1503 ATTCCCACTGATAAGCGTACTTATAGTAAGGAGTGTCTACACCGAGGAGACATA 1562
Db 1715 CGTGCCTGCAGAGGAGGCTTCTATAGTCAAGAGAGATTCTCGCTACAGAGACATA 1774
QY 1563 TCTGAAGGATCTGAAAGTATATCTCTGTTTTCAGAGCAGCAGTGAAGAGAGCGC 1622
Db 1775 CTTCAAGGATTTAAGAGTATTACCTGTGTTCCGAGCGCAGTGGTGAAGGAGCGC 1834
QY 1623 CATGCCGAGAGCAGTGAAGAGTCTATATCCCGAATTTTGAACCTTTGACAAATTTCA 1682
Db 1835 CATGCTCGGACCTGATGAGCTGTCTTCTCAACATCGATCCCTATGAGTTCCA 1894
QY 1683 TACTAATTTCTCAGGAATTTGAGCAAGCAGCTTGGCTGTGGGAGCGGCTCAATGC 1742
Db 1895 CAGAGGCTTCTCGGAGGTGAGCAGAGGCTGGCAGTCTGGGAGGCGCTCCAAAGC 1954
QY 1743 CCAAAATCAGAG---ATTACCAAGAATCGCGAGTGTGCTGCTCAAGAACATTC 1792
Db 1955 CCACAAAAGGAGCAGTATCAAGAAATCGGGGACATCTGCTCAGGAACATGC 2007

RESULT 6
BC004009 2632 bp mRNA linear ROD 07-AUG-2002
LOCUS BC004009
DEFINITION Mus musculus, clone IMAGE:3493093, mRNA, partial cds.

ACCESSION BC004009
VERSION BC004009.1
KEYWORDS GI:13278387
SOURCE house mouse,
ORGANISM Mus musculus

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 8 Row: p Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
Source

1..2632 Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZCH II"
/clone="IMAGE:3493093"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..1263
/codon_start=1
/product="Unknown (protein for IMAGE:3493093)"
/protein_id="AAH04009.1"
/db_xref="GI:13278387"
/translation="VNLKNIQGMKHLAAHLKHSLEALETSIKGRRLLEHFCRDFE
LOKCYLPLNTFLRLPLHRLMXYKHVLEBLCKHHPNADFRCDRAALAEITFEMVAOL
HGTMLKMFQKLEKLDLIGIDNLVLPREFIRGLSKLSKGLQLORMFLFNDV
LUTSRGLTASNOFKVHGPOLPLYGMTIEESEEWGVPHCLTIRGQRQSLIVASRSSE
MEKMDIQMADLAEKSNPTPELLASPPDNKSPDEATQDESDLSASRTSLE
RQAPHRGTMVHVCWHRSTSVSVDFTAVENQLSGLNLRKFNKSNGLKLVVFTNF
CLFFYKHQDSHPLASLPLGLSLTIPSENIHKDYVFKLHFKSHVYVFAESEVTF
ERWMEVIRSATSSASRAHLHSHKSHLY"

CDS

BASE COUNT 632 a 720 c 671 g 609 t
ORIGIN

Query Match 20.4%; Score 478.8; DB 10; Length 2632;
Best Local Similarity 89.3%; Pred. No. 1.2e-104;
Matches 516; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1774 GTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGCTCACCTGTGGAAGACAGC 1833
Db 1 GTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGCTCACCTGTGGAAGACAGC 60

QY 1834 GAGGCTTGGAGCCCTGGAGAAATGAATCAAGAGCTCCCGGCGCTGGAGAACTTCTGC 1893
Db 61 GAGGCTTGGAGCCCTGGAGAAATGAATCAAGAGCTCCCGGCGCTGGAGAACTTCTGC 120

QY 1894 AGAGACTTTGAGCTGCAGAGGTGTGTTACCTACCGCTCAACACCTTCCTCTCGGGCCA 1953
Db 121 CGAGACTTCGAGCTGCAGAGGTGTGTTACCTACCGCTCAACACCTTCCTCTCGGGCCA 180
QY 1954 CTGACCGGCTCATGACACTACAGCAGGTCTCTGGAGCGCTGTGCAACACACCGCCGC 2013
Db 181 CTGACCGGCTCATGACACTATAAGCATGTCTCTGGAGCGCTGTGCAACACACCGCCACA 240
QY 2014 AGCCACCGGCTCTAGGACTCCGCGCGCTTTGGCAGAGATCACGAGATGTTGGCA 2073
Db 241 AACACCGGCTCTAGGACTCGAGACTCGAGCTCGCTGGCGAGATCACAGAGATGTTGGCC 300
QY 2074 CAGCTCCACGGTACGATGATCAAGATGAGAAATTTCCAGAACTGCACGAACTCAAGAAA 2133
Db 301 CAGCTGCACGGGACCATGATCAAGATGAGAACTTTCCAGAACTGCATGAGCTCAAGAAA 360
QY 2134 GATTTGATGGCATTCACAACTCTTGTTGCTCGGGAAGGAGTTTCATCGCTGGGCAGC 2193
Db 361 GATCTGATCGGCATTCACAACTCTTGTTGCTCGGGAAGGAGTTTCATCGCTGGGCAGC 420
QY 2194 CTGACCAAGCTCTCGGGAAGGCTCCAGCAGCGCATGTTCTCTGTTCAACGAGCTC 2253
Db 421 CTCAGCAAGCTCTCGGGAAGGCTTCAGCAGCGCATGTTCTCTGTTCAACGATGTC 480
QY 2254 CTGCTATACAGCGCGGGGCTGACGGCTCCAACTCAATCACTTTAAAGTCCACGGGCGCTC 2313
Db 481 TTGCTGTATACCAAGCGGGGCTGACAGCATCTAATCACTTTAAAGTCCACGGGCGCTC 540
QY 2314 CCCTCTATGGCATGACGATTGAGGAGCGGAGAGCA 2351
Db 541 CCATCTATGGCATGACGATCGAGGAGGAGTGGAGGA 578

RESULT 7
HSM802608 2888 bp mRNA linear PRI 23-MAR-2000
LOCUS Homo sapiens mRNA; cDNA DKFp762P046 (from clone DKFp762P046).
DEFINITION AL161984
ACCESSION AL161984.1 GI:7328120
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2888)
AUTHORS Bloeker,H., Boecker,M., Brandt,P., Mewes,H.W., Weil,B. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFp762P046) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
FEATURES
source location/Qualifiers
1..2888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="DKFp762P046"
/tissue_type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal 2852..2857
polyA_site 2859
BASE COUNT 879 a 592 c 620 g 797 t

ORIGIN
Query Match 14.6%; Score 342.8; DB 9; Length 2888;
Best Local Similarity 68.2%; Pred No. 8.9e-72;
Matches 476; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 295 AGGTACCTGTTCCGGCTGCAGGTGAAGCAGGACTTTGGCTCAAGGCGAGTTTGACGTGTAAT 354
Db 123 AGGTATCTTTTACTCTTCAAAATAAAGAGGATTTGGCTCTAGGAAGGCTTCCATGCAGT 182
QY 355 GACACAGCGCAGCTCTCTTTGATTTTACACATCTGCAATCTGCAATCTGGGATTTTGTAT 414
Db 183 GACAACCTGTACAGCGTTGATGTATCTCACATCTTTACAAATCAGAACTTGAGACTTTTCAT 242
QY 415 GAAGCCTTGGCAGAGAGCAGCTTAGCAAAAATAAATACATACCTCAGCAACGCGCACTA 474
Db 243 GAAGAACAAGATAGGAAGCATCTGGCAAACTCGGTACTTTACCAAAACCAAGACTGTTTA 302
QY 475 GAGGACAAATCGTGGAAATTTCCACCATACCCATTTGGCAAAACACACAGCAGAAATCAGAT 534
Db 303 GAGGCAAGATCATGCACCTTTCATCAGAAGCACATTTGGCAGGAGCCAGCTGAATCTGAC 362
QY 535 TTCCAGCTCCTAGAGATTTGCCGCTGGCTAGAGATGTATGGAATCGGGTTGACCCGGCC 594
Db 363 ATTCTGCTACTGGACATAGCAAGAGCTGGATATGTATGGCATCAGGCTCACCCGCC 422
QY 595 AAGGACAGGGAAGGACGAAGATCAATCTGGCCGTTGCCAACACGCGGAATCTAGTGTATT 654
Db 423 AGTGATGTTGAAGGATGACAGATTCACCTGGCTGTTGTCATCATGGGATGCTGGTGTTA 482
QY 655 CAGGGTTTCACTTAAGTCAATGCTTCACTGGGCAAGTTCGGAAGCTGAGCTTCAAG 714
Db 543 AGAAAGCATTTTCTCATCAAACTTTCATGCAATATCTTGGTGTGTGCAAGGATACCTTG 602
QY 775 GAATTCCTGATGCCAGTCGGGATTTCTCAAGTCTCTCTGGAATACTCTGTGAACAT 834
Db 603 GAGTTCACCATGCCAGCCGAGATGCTTCAAGGCTTTCTGGAAGACTTGTGTGAATAC 662
QY 835 CATGCTCTTTAGACTTTTGAAGAGCCCAACCAAGCCCAAGCCCTCTCTTTAGC 894
Db 663 CATGCTTTCTCAGGCTTTTCGGAAGAGCCCAAAATCAAAAGCCCAAAAGCCCTACTGCA 722
QY 895 CGGGGTATCATTTTCGGTTCAGTGGTGGACTCAGAACGAGGTTCTCGACTATGTTAA 954
Db 723 AAGGTTTCAGTTTCCGCTATAGTGACGACCAACCCAAAGGCAACTTTTGAATATGGGAGA 782
QY 955 GAAGGAGGACATAAAGAGGTGCAGTTTGAAGGGAAGCA 992
Db 783 AAGGAGGCTCAAGAGCTTGCCATTTGAAGGGAANCA 820

RESULT 8
BC027077 2431 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone IMAGE:5010682, mRNA, partial cds.
DEFINITION BC027077
ACCESSION BC027077.1 GI:20071584
VERSION
KEYWORDS house mouse.
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2431)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kovis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 45 Row: c Column: 12
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES
 source

Location/Qualifiers
 1. .2431
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N-3"
 /clone="IMAGE:5010682"
 /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
 old mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam2"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"
 <1. .1065
 /codon_start=1
 /product="Unknown (protein for IMAGE:5010682)"
 /protein_id="AA27077.1"
 /db_xref="GI:20071585"
 /translation="YKHVLRLCKHPPNHADFRDRAALAEITEMVAQLHGTMIKME
 NFQKHELKDLIGLNLVYPPREFIRLGSLSKLGLOORFFFLFNDVLLXTSRGL
 TMAQIDKVGQLPYGWTIESEEWGVPHCLTLRGQRQSIIVAASRSRSMEXKMEID
 QASNDLAERKSNPDTPELLASSPDKPDATAODESDIDLSASRTSLERQAPHRGNI
 TMVYCVHBRSTSVSMVDFSAVENQLSGNLLRKFKNNGHQLWVFTNCLFFYKSH
 QDSHPASLPLLLSHYTSISESNIHKDIYVFLHFKSHVIYFRAESEYTFERWMEVIR
 SATVSSAAHLLSHKRESHLY"

CDS

BASE COUNT	581 a	659 c	614 g	577 t
ORIGIN				

Query Match	13.3%	Score 312.8;	DB 10;	Length 2431;
Best Local Similarity	88.9%	Pred. No. 1.6e-64;		
Matches 336;	Conservative	0;	Mismatches 42;	Indels 0;
Gaps	0;			

QY 1972 TACAAGCAGGTCTCTGGAGCGGCTGTGCAACACACACCGCGGAGCCAGCTTCAGG 2031
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 TATAAGCATGCTCTGGAGAGGCTGTGCAGGACACACCCACCAACACCGCCGCACTTCAGG 60
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2032 GACTGCGAGCCGCTTTGGCAGAGATCACGAGATGGTGGCACAGCTCCACGGTACGATG 2091
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 61 GACTGCAAGCTCGGCTGGCGGAGATCACAGATGTTGGCACAGCTGCATGTTACCATG 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2092 ATCAAGATGAGATTTCCAGAAGCTGCAGACTCAAGAAAGATTTGATTTGGCATTTGAC 2151
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 121 ATCAAGATGAGAACTTCCAGAAGCTGCATGAGCTCAAGAAAGATCTGATCGGCATTGAC 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2152 AATCTTGTGTTCGGGAAGGAGTTTCATCCGTCGTGGCAGCCTCAGCAAGCTCTCGGG 2211
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 181 AATCTTGTGACCCGAAGGAGTTTCATCCGCTGGGAGCCTCAGCAAGCTCTCGGG 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2212 AAGGGGCTCCAGACGCGCATGTTCTTCTGTTCACACAGCGTCCTGCCTATACACAGCGG 2271
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 241 AAGGGGCTTACGACGCGCATGTTCTTCTGTTCACCATGTTCTTCTGTATACACGCGG 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2272 GGCGTGACGGCCTCCAACTAGTTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGACG 2331
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db	301	GGACTGAGCGATCTAATCAGTTTAAAGTCACGAGACGCTCCCACTCTATTGCATGACG	360
Qy	2332	ATTGAGGAGCGGAAGACGA	2351
		II IIIIIIII II II II	
Db	361	ATCGAGGAGAGTGAGGAGGA	380
RESULT 9			
AL161896			
LOCUS			
DEFINITION	ALI61896 96183 bp DNA linear PRI 20-JAN-2001 Human DNA sequence from clone RP11-261P24 on chromosome 13, complete sequence.		
ACCESSION	ALI61896		
VERSION	ALI61896.16 GI:12330752		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
1 (bases 1 to 96183)			
Phillimore,B.			
Direct Submission			
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,			
CBL01 SA, UK. E-mail enquiries: humquery@sanger.ac.uk			
Requests: clonerequest@sanger.ac.uk			
On Jan 22, 2001 this sequence version replaced gi:12329254.			
During sequence assembly data is compared from overlapping clones.			
Where differences are found these are annotated as variations			
together with a note of the overlapping clone name. Note that the			
variation annotation may not be found in the sequence submission			
corresponding to the overlapping clone, as we submit sequences with			
only a small overlap as described above.			
This sequence has been finished according to sequence map criteria			
as follows. An attempt is made to resolve all sequencing problems,			
such as compressions and repeats, but not necessarily within known			
annotated repeat sequence elements. Where the sequence is			
ambiguous, there is an annotation using the 'unsure' feature key.			
The following abbreviations are used to associate primary accession			
numbers given in the feature table with their source databases:			
Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information			
on the WORMPEP database can be found at			
http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence			
was generated from part of bacterial clone contigs of human			
chromosome 13, constructed by the Sanger Centre Chromosome 13			
Mapping Group. Further information can be found at			
http://www.sanger.ac.uk/HGP/Chr13			
RP11-261P24 is from the library RPCI-11.1 constructed by the group			
of Pieter de Jong. For further details see			
http://www.chori.org/bacpac/home.htm			
VECTOR: pBACE3.6			
IMPORTANT: This sequence is not the entire insert of clone			
RP11-261P24 It may be shorter because we sequence overlapping			
sections only once, except for a 100 base overlap.			
The true right end of clone RP11-261P24 is at 96183 in this			
sequence. The true left end of clone RP11-31K22 is at 59428 in this			
sequence. The true right end of clone RP11-573N10 is at 100 in this			
sequence.			
FEATURES			
source	Location/Qualifiers		
	1..96183		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="13"		
	/clone="RP11-261P24"		
	/clone_lib="RPCI-11.1"		
repeat_region	1301..1335		
	/note="SN5 repeat: matches 34..68 of consensus"		
repeat_region	1432..1506		
	/note="MER20 repeat: matches 1..75 of consensus"		
repeat_region	3518..3854		
	/note="LIMB8 repeat: matches 5837..6173 of consensus"		
repeat_region	5528..5599		
	/note="12 copies 6 mer tatta 69% conserved"		
repeat_region	5533..5604		
	/note="12 copies 6 mer tatta 69% conserved"		


```

|||||
Db 95725 AGCTCCGCGCAGGAAAGCAACGAAGTTTCCTCCGCGGGAGCGGGTTCGCACCCGAGC 95784
|||||
QY 1180 CTTGCGCGCAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCCCTCGCGGCC 1239
|||||
Db 95785 CTTGCGCGCAGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCCCTCGCGGCC 95844
|||||
QY 1240 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1299
|||||
Db 95845 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 95904
|||||
QY 1300 CAGCCAAGCACAGG 1313
|||||
Db 95905 CAGCCAAGCACAGG 95918
|||||

RESULT 10
AL137249 Human DNA sequence from clone RP11-111L24 on chromosome
DEFINITION 13q31.3-32.3, complete sequence.
ACCESSION AL137249
VERSION AL137249.29 GI:18655955
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 106578)
Garner,P.
Direct Submission
Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced gi:18121451.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-111L24 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-111L24. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-295817 is at 104579 in this
sequence. The true right end of clone RP11-261P24 is at 2000 in
this sequence.
FEATURES
source location/Qualifiers
1. 106578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="q31.3-32.3"

```

```

/misc_feature
/clone="RP11-111L24"
/clone_lib="RPCI-11.1"
79533
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 800bp by restriction digest data"
BASE COUNT 28669 a 25979 c 25140 g 26790 t
ORIGIN

Query Match 10.8% Score 254; DB 9; Length 106578;
Best Local Similarity 100.0%; Pred. No. 3.5e-50;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1060 CAGTCTCAGCAGAGCACCAGCCTTACATTGAGAGGTGCGCAATCTCCAGGGGCCAG 1119
|||||
Db 1482 CAGTCTCAGCAGAGCACCAGCCTTACATTGAGAGGTGCGCAATCTCCAGGGGCCAG 1541
|||||
QY 1120 AGCTCCGCGCAGGAAAGAACCGAAGTTTCCTCCGCGGGAGCGGGTTCGCACCCGAGC 1179
|||||
Db 1542 AGCTCCGCGCAGGAAAGAACCGAAGTTTCCTCCGCGGGAGCGGGTTCGCACCCGAGC 1601
|||||
QY 1180 CTTGCGCGCAGGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCCCTCGCGGCC 1239
|||||
Db 1602 CTTGCGCGCAGGAGAGAGCCCGCGGGTAAACAGAGCGGAGCGGAGCGCCCTCGCGGCC 1661
|||||
QY 1240 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1299
|||||
Db 1662 ACGGAGGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGG 1721
|||||
QY 1300 CAGCCAAGCACAGG 1313
|||||
Db 1722 CAGCCAAGCACAGG 1735
|||||

RESULT 11
AC022669/c AC022669 152053 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT
DEFINITION SEQUENCE, 33 unordered pieces.
ACCESSION AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE 2 (bases 1 to 152053)
Homo sapiens chromosome 13, clone RP11-96B23
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
ROY,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 152053)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,

```

Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Lakoczek, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenca, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4938

Center clone name: 96_B_23

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads

Sequencing vector: Plasmid; n/a; 0.0% of reads

7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136451 bases at least Q40

Consensus quality: 143612 bases at least Q30

Consensus quality: 146365 bases at least Q20

Insert size: 158000; agarose-fp

Insert size: 148853; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.

NOTE: this is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 784: contig of 784 bp in length

785 884: gap of 100 bp

885 1956: contig of 1072 bp in length

1957 2056: gap of 100 bp

2057 3113: contig of 1057 bp in length

3114 3213: gap of 100 bp

3214 4817: contig of 1604 bp in length

4818 4917: gap of 100 bp

4918 6582: contig of 1665 bp in length

6583 6682: gap of 100 bp

6683 8059: contig of 1377 bp in length

8060 8159: gap of 100 bp

8160 22369: contig of 14210 bp in length

22370 22469: gap of 100 bp

22470 24974: contig of 2505 bp in length

24975 25074: gap of 100 bp

25075 27206: contig of 2132 bp in length

27207 27306: gap of 100 bp

27307 29420: contig of 2114 bp in length
29421 29520: gap of 100 bp
29521 32000: contig of 2480 bp in length
32001 32100: gap of 100 bp
32101 36330: contig of 4230 bp in length
36331 36430: gap of 100 bp
36431 39067: contig of 2637 bp in length
39068 39167: gap of 100 bp
39168 42564: contig of 3397 bp in length
42565 42664: gap of 100 bp
42665 46664: contig of 4000 bp in length
46665 46764: gap of 100 bp
46765 51047: contig of 4283 bp in length
51048 51147: gap of 100 bp
51148 55269: contig of 4122 bp in length
55270 55369: gap of 100 bp
55370 58831: contig of 3462 bp in length
58832 58931: gap of 100 bp
58932 64077: contig of 5146 bp in length
64078 64177: gap of 100 bp
64178 68808: contig of 4631 bp in length
68809 68908: gap of 100 bp
68909 74170: contig of 5262 bp in length
74171 74270: gap of 100 bp
74271 79296: contig of 5026 bp in length
79297 79396: gap of 100 bp
79397 84724: contig of 5328 bp in length
84725 84824: gap of 100 bp
84825 89713: contig of 4889 bp in length
89714 89813: gap of 100 bp
89814 94721: contig of 4908 bp in length
94722 94821: gap of 100 bp
94822 100196: contig of 5375 bp in length
100197 100296: gap of 100 bp
100297 107015: contig of 6719 bp in length
107016 107115: gap of 100 bp
107116 114507: contig of 7392 bp in length
114508 114607: gap of 100 bp
114608 122671: contig of 8064 bp in length
122672 122771: gap of 100 bp
122772 128052: contig of 5281 bp in length
128053 128152: gap of 100 bp
128153 138583: contig of 10431 bp in length
138584 138683: gap of 100 bp
138684 150248: contig of 11565 bp in length
150249 150348: gap of 100 bp
150349 152053: contig of 1705 bp in length.

FEATURES

source

1. .152053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13"
/clone="RP11-96B23"
/clone_lib="RPC1-11 Human Male BAC"
1. .784
/note="assembly_fragment"
vector_side:left"

misc_feature

885. .1956
/note="assembly_fragment"
2057. .3113
/note="assembly_fragment"
3214. 4817
/note="assembly_fragment"
4918. .6582
/note="assembly_fragment"
6683. .8059
/note="assembly_fragment"
8160. 22369
/note="assembly_fragment"
22470. .24974
/note="assembly_fragment"

QY	472	CTAGAGGACAAAATCGTGGAAATTTCCACCAATACCAATTTGACAAAACACCAGCAGAAATCA	531
Db	1665	ATGCAGCGAAAAGATTATGGAGAACCAAAAAGCATGTGGTCAATCCCGCGGAGGCG	1724
QY	532	GATTTCCAGCTCTAGAGATTGCCGTCGCTAGAGATGATGAATCCGGTTGCACCCG	591
Db	1725	GACCTTAACCTCTTTGGAGACGGCTCGGAGGTGTGAGCTGTATGCGATGAATAATGCATCCG	1784
QY	592	GCCAAGGACAGGGAAGCAGCAAGATCAATCTGCCGCTGCAACACACGGGAATTCCTAGTG	651
Db	1785	GCGAAGGATGTGGAGGGGTGCCGCTTAATCTGCTGTGCCACATCGGCATCACAGTC	1844
QY	652	TTTCAGGGTTTCACTAAGATCAATGCGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTC	711
Db	1845	TTCCAGAACATCAGCGGGATCAACACCTTCTCGTGGGTAAAGATACGCAAGATTTCTTTC	1904
QY	712	AAGGAAGGCGCTTCTCATCAAGCTCGGCCAGATGCCAATAGTGGTACCAGATACC	771
Db	1905	AAGCGCAAGCGATTCTCTGGTCAAACTGCATCCAGAGGATATGATATATCAAGAGATACC	1964
QY	772	TTGGAATTCCTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGAAAATCTGTGTGTGAA	831
Db	1965	GTGAGTCTCTTCGAGGGTGCACAGTGCAAAAATCTCTGAAAAAATGCGTCGAA	2024
QY	832	CATCATGCCCTTTTAGACTTTTGTGAAGGCCAA---ACCAAGGCCAACGCCGTCCTC	888
Db	2025	AATCAGCGATTCTCCGATGCATCGCGTACAAAATACGCCAGCGCAAAATCGCGTT	2084
QY	889	TTTAGCCGGGGTCATCATTTTCGTTTCAGTGGTCCGACTCAGAAGCAGGTCTCGACTAT	948
Db	2085	CTCTCGGGGTAGTTCATTCCGCTATAGCGGAAAAACCCAGAAGCATATTCAGTTTC	2144
QY	949	GTTAAAGAGGAGGACATAAGAAGGTGCAGTTTCAAGAGGAAGCACACAA	998
Db	2145	GTTCGCGAAATATGTGAAGCGCCAAACTTCCAAAGGTATAAACGAA	2194
RESULT	13		
AC094777			
LOCUS			
DEFINITION	AC094777	224310 bp DNA linear	HTG 24-AUG-2002
		Rattus norvegicus clone CH230-4N8, *** SEQUENCING IN PROGRESS ***,	
ACCESSION	AC094777	65 unordered pieces.	
VERSION	AC094777.6	GI:22164997	
KEYWORDS	HTG: HTGS PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
	1 (bases 1 to 224310)		
	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,		
	Allen, C., Allen, H., Altschroeder, S., Amin, A., Anguiano, D.,		
	Anylebeche, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,		
	Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,		
	Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,		
	Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,		
	Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,		
	Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,		
	Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,		
	Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,		
	Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,		
	Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,		
	Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,		
	Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,		
	Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,		
	Georgiegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,		
	Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,		
	Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,		
	Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,		
	Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,		
	Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,		

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowls, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhsawa, L., Loulsegad, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morris, N., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G.,
Olarunsaogun, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
Reuter, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
Rever, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E.,
Song, X.-Z., Sorelle, R., Sosa, J.C., Stelmie, M., Strong, R., Sutton, A.,
Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,
Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, V., Yoon, Y., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 224310)
Worley, K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 224310)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 9, 2002 this sequence version replaced gi:22094284.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GBIZ
Center clone name: CH230-4N8

Summary Statistics
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 172428 bases at least Q40
Consensus quality: 179331 bases at least Q30
Consensus quality: 184096 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1011: contig of 1011 bp in length
* 1012 1111: gap of unknown length
* 1112 2458: contig of 1347 bp in length
* 2559 2558: gap of unknown length
* 2559 3836: contig of 1278 bp in length
* 3837 3936: gap of unknown length

COMMENT On Jun 23, 2002 this sequence version replaced gi:21218511.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA0090H14
----- Summary Statistics -----

Sequencing vector: M13; 0%
Chemistry: Dye-terminator ET; 100%
Assembly program: Phrap; version 0.990319
Consensus quality: 216976 bases at least Q40
Consensus quality: 217754 bases at least Q30
Consensus quality: 218341 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 221868; sum-of-contigs
Quality coverage: 15.60 in Q20 bases; agarose-fp
Quality coverage: 12.41 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 8441: contig of 8441 bp in length
* 8542: gap of unknown length
* 27277: contig of 18736 bp in length
* 27378: gap of unknown length
* 42339: contig of 14962 bp in length
* 42340: gap of unknown length
* 42440: contig of 29099 bp in length
* 71539: gap of unknown length
* 71639: gap of unknown length
* 99587: contig of 27949 bp in length
* 99588: gap of unknown length
* 134048: contig of 34361 bp in length
* 134149: gap of unknown length
* 221350: contig of 87201 bp in length
* 221450: gap of unknown length
* 221502: contig of 53 bp in length.

FEATURES

source
1. .221502
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-90H14"
1. .8441
/note="assembly_name:Contig59"
8542. .27277
/note="assembly_name:Contig60"
27378. .42339
/note="assembly_name:Contig61"
42440. .71538
/note="assembly_name:Contig62"
71639. .99587
/note="assembly_name:Contig63"
99688. .134048
/note="assembly_name:Contig64"
134149. .221349
/note="assembly_name:Contig65"
221450. .221502
/note="assembly_name:Contig19"
61405 a 54522 c 50450 g 54373 t 752 others

Query Match

8.5%; Score 199.4; DB 2; Length 221502;

Best Local Similarity 89.2%; Pred. No. 5.8e-37;
Matches 215; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 1801 AAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGAGGCGCTTGGAGGCGCTTGGAGAATGGA 1860
|||||
Db 74064 AAGCACTGGGGCCACCTGTGGAAGCACAGTACAGGCGCTTGGAGGCGCTTGGAGACCTCC 74123
|||||
QY 1861 ATCAAGAGCTCCCGCGGCTGGAGAAGTCTTTCGAGAGAGCTTTGAGCTGCAGAAAGTGTGT 1920
|||||
Db 74124 ATCAAGGGCTCGCGGCGCTGGAACATTTCTGCCGAGACTTCGAGCTGCAGAAAGTGTGT 74183
|||||
QY 1921 TACCTACCGCTCAACACCTTCCTCTCGGCGGCTGACCGGCTCATGCATCAAGCAG 1980
|||||
Db 74184 TACCTGCGCTCAACACCTTCCTCTCGGCGGCTGACCGGCTCATGCATCAAGCAT 74243
|||||
QY 1981 GTCCTGGAGCGGCTGTGCAACACACCGCGCGGAGCAGCCGCTTCAGGAGCTTGCAGG 2040
|||||
Db 74244 GTCCTGGAGCGGCTGTGCAACACACCGCGCGGAGCAGCCGCTTCAGGAGCTGCAGA 74303
|||||
QY 2041 G 2041
Db 74304 G 74304

RESULT 15

AF156225
LOCUS Homo sapiens protein 4.1 (EPB4.1) mRNA, complete cds.
DEFINITION
ACCESSION AF156225
VERSION AF156225.1 GI:5353737
KEYWORDS
SOURCE Homo sapiens.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Huang,S.C., Wang,C., Lichtenauer,U., Vortmeyer,A. and Zhuang,Z.
TITLE Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2595)
AUTHORS Huang,S.C., Wang,C., Lichtenauer,U., Vortmeyer,A. and Zhuang,Z.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1999) NCI, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA

FEATURES

Location/Qualifiers
1. .2595
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="lp36"
/cell_line="LAN5"
/cell_type="neuroblastoma"
1. .2595
/gene="EPB4.1"
1. .2595
/gene="EPB4.1"
/codon_start=1
/product="protein 4.1"
/protein_id="A04222.1"
/db_xref="GI:5353738"

gene

CDS

/translation="MTTEKSLVTEAENSHQHQKEEGEAINGQQEQEQEESCQTAAE
GDNWCEHKLSKASNGDTTHEDLTNKRERTSENRLSRLFSFLKRPKSQVSEEGKEV
ESDKEGEGGQKEIEFGTSLDEEILKAPAAPELTDPLDLHLSAETAPQAE
ELREDPDFEIKEGGLEECSEKIEVKESPOSKAETELKASOKPIRKRNHCKVSLID
DTVYECVVEKHAKGQDILLKRVCEHNLLEDEYFGLATWDNTSKTLDLSAKEIKQVR
GVPWNFTFNKVPDPDAQLETITRYLCLQLRODIVAGRLPSPFATLALGSIYTIQ
SELGDYDPELHGVDIYDSDFLAPNQTKLEEKVMELHKSRKMTTPAQADLEFLNARK
LSMYGVDLHKAKDLEGVDIILGVCGSLLYKDLRLNRPFPKVLISLYKRSFFIK
IRPGEQYESTIGFLPSRAAKKLWKCVEHHTFFRLTSTDTPKSKFTALGSKFR
VSGTQATQASALIDRPAPHFRTASKRASRLSDGAAAVDSADRPRTSAIPAIT
GOVAEGGLDASAKTVPYKQKFTYKAEVKDEDEPPEQAEPTPEAKVKETKHEVT
VPTNSNGQOTKLAETEDLIMRKKRERLDGENIYIRKHSMLLEDLDKSOEELEKHH
ASISELKKNFMESVPEPRPSEWDRRLSTHSPFTLININGQIPTGEPPLVATQIVTIS

DNANAVKBIPIKDVPIVHTETKITYEAAQTDDNSDLDPGVLLTAQTIITSETPSST
TTTQITKTVKGGISITREKIRIVITGDADIDHDQVLVQAIKEAKEQHPDMSVTKVVVH
QTEIADE*

Search completed: December 3, 2002, 20:46:10
Job time : 5413.44 secs

BASE COUNT 881 a 563 c 601 g 550 t
ORIGIN

Query Match 8.3%; Score 194; DB 9; Length 2595;

Best Local Similarity 52.9%; Pred. No. 8e-36;

Matches 491; Conservative 0; Mismatches 425; Indels 12; Gaps 3;

```
QY 15 CGTCTCCATCCAGATCGTGATGACACCCAGGAGGATTTGAAGTTCACCAAG 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 CATGCACTGCAAGTTCTTTGTTGGATGACACAGTTTATGAATGTTGTGGAGAAACA 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 AGCTCTGGGAAGTCTGCTGGATGACAGTTTGCACACCTCAACCTCGTGGAGGTGA 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 687 TGCTAAGGGACAGATTGCTTAAACGAGTATGTGAGCATCTCAATCTTTGGAGAAGA 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 135 CTATTTTGGCTCGAGTTTCTCTGATCACAAGAAAGATCAGGGTGTGGCTGGATCTCTAAA 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 CTATTTTGGCTAGCCATTGGGATAACACACCTCTAAGACATGGCTGGATTCGCGCAA 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 ACCATTTGTGAACAGATTAGAGGCCAAGCAGCGTTGTTGTTAACTTTGTGTTGAAATT 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 AGNAATAAAAAGCAGGTTCGTGG---TGTCCCTTGGAAATTTTACATTTTAATGTAAGTT 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 CTTTCGCCCTGACACACACAACTCCAAGAAAGAACTACAAAGGTACCTGTTCGCGTGA 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 TTATCCACCTGACCCAGCAGGTTAACAGAGACATAACAAGATATATTATGCTCTTCA 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 GGTGAAGCAGGACTTGGCTCAAGGCAGGTTGAGTGTATGACACAGCGCAGCTCTCTT 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 924 GCTTCGGCAGGACATAGTTGCAGGAGCTGTGCCCTGTCTTGTGCAACCTTAGCATATT 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 375 GATTTCCACATTTGCTCAATCTGAGATTTGGGATTTTGTAGAGCCCTTGGACAGAGCA 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 984 AGTTCTTACACCATCCAGCTGGAAGTGGAGACTACAGCCCAAGAACTCCATGGCGTGA 1043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 CTTAGCAAAAATAAATAGATACC-----TCAGCAAGACGCACTAGAGGACAAAAATCGT 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1044 TTATGTTAGTGATTTTAAACTGGCCCCGAATCAGACCAAGGAAGTTGAAGAGAGGTGAT 1103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 489 GGAATTTACACATAACCAATTGGGCAAAACACAGCAGAGAAATCAGATTTCCAGCTCTAGA 548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1104 GGAACCTGCATAGTCATAGAGTCCATGACTCCAGCTCAGGCTGACTTGGAGTTTCTTGA 1163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 549 GATTGCCGCTGCGTAGAGATGATGGAATCCGGTTGACCCCGCCCAAGGACAGGGAAG 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1164 GAATGCCAAAAGTTGCTATGATGGAGTTGATCTTCATAAGCAAGGAGCTTGGGAAG 1223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 609 CACGAAGATCAATCTGGCGGTTGCCAACACGGAATTTAGTGTTCAGGGTTTCACTAA 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1224 AGTAGATATCATCTAGTGTCTGTCTAGTGGCTTCTGGTTTACAAGATAAGCTGAG 1283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 669 GATCAATGCTTCACTGGGCCAAGGTGGGGAAGCTGAGCTTCAAGAGGAAGCGTTTCT 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1284 AATTAAACCGCTTCCCTTGGCCCAAGTGTGAAGATTTCTTATAAACCTAGTAGCTTTT 1343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 729 CATCAAGCTCCGGCAGATGCCAATAGTGGTACCAAGATACCTTGGAAATTCCTGATGGC 788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1344 CATCAAGATTCGGCCCTGGAGCAAGAGCAGTATGAAGTACCATCGGATTCAAACTTCC 1403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 789 CAGTCGGGATTTCTGCAAGTCTTCTGGAAAATCTGTGTTGAACATCATGCCCTTCTTTAG 848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1404 CAGTTACCGAGCAGCTAAGAAATATGGAAGTCTGTGTAGAACATCACACGTTTTTTCAG 1463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 849 ACTTTTGAAGCCCCAACCAAGCCCGCTCTTTAGCCGGGGGTCAATCAT 908
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1464 A---TTGACATCTACAGACACCATTCCTCCAAAAGCAAAATTTCTTGGCTAGGATCCAAAT 1520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 909 TCGGTTCACTGTGGGACTCAGAAGCAG 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1521 TCGATACAGTGGCGGACTCAAGCTCAG 1548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: December 3, 2002, 14:48:04 ; Search time 371.636 Seconds
(without alignments)
14246.337 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_2501
Perfect score: 2351
Sequence: 1 ccttcaggaaactcgtgtc.....attgaggagcgaagacga 2351

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*									
		1:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*						
		2:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*						
		3:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*						
		4:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*						
		5:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*						
		6:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*						
		7:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*						
		8:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*						
		9:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*						
		10:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*						
		11:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*						
		12:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*						
		13:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*						
		14:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*						
		15:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*						
		16:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*						
		17:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*						
		18:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*						
		19:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*						
		20:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*						
		21:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*						
		22:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*						
		23:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*						
		24:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*						
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES									
8									
Result	Query	Score	Match	Length	DB	ID	Description		

1	2351	100.0	3442	20	AAX79183	Human chondrocyte-			
2	2349.4	99.9	4687	21	AAA08582	Human cytoskeleton			
3	2335.2	99.3	3187	23	AAS64760	DNA encoding novel			
4	762	32.4	3094	21	AAC98992	Human pancreatic c			
C	5	254	10.8	1824	23	AAS64761	DNA encoding novel		
	6	228	9.7	1220	23	ABL02335	Drosophila melanog		
c	7	223.6	9.5	241	21	AAA42150	Human secreted exp		
	8	192.4	8.2	4922	23	ABV30214	Human prostate exp		
9	181.8	7.7	3643	24	ABK84074	Human cDNA differe			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

10	180.4	7.7	6263	24	ABK70019	Human cDNA for the
11	171.4	7.3	2916	23	ABL12469	Drosophila melanog
12	171.4	7.3	3470	23	ABL18111	Drosophila melanog
13	171.4	7.3	5897	23	ABL18211	Drosophila melanog
14	169.6	7.2	3620	21	AAA08581	Human cytoskeleton
15	162.6	6.9	14917	24	AAS23338	Human lung specifi
16	154.6	6.6	426	23	AAS58814	cDNA #1490 encodin
17	144	6.1	3984	18	AAT58627	protein tyrosine p
18	144	6.1	3984	18	AAT58627	protein tyrosine p
19	144	6.1	3984	24	AAX78463	Human PTPH1 CDNA.
20	142.8	6.1	2872	22	AAS08887	PTPH1 CDNA. Homo
21	142.8	6.1	3313	23	ABN93400	Human protein tyro
22	138.2	5.9	3106	23	ABN93399	Human gene GS91552
23	134.2	5.7	3031	23	ABV22627	Human gene GS91552
24	134.2	5.7	3031	23	ABV22627	Human prostate exp
25	133.6	5.7	2067	24	ABN85314	Human cytoskeleton
26	133.6	5.7	3166	22	AAH17787	Human cDNA sequenc
27	133.4	5.7	4356	23	ABL29969	Drosophila melanog
28	132.2	5.6	2846	24	ABL60869	Human cytoskeleton
29	121.8	5.2	2337	24	ABQ73378	Human red blood ce
30	106	4.5	1718	22	AAS27016	cDNA encoding nove
31	105.8	4.5	1747	23	ABL03867	Drosophila melanog
32	105.8	4.5	4150	23	ABL03866	Drosophila melanog
33	105.2	4.5	716	22	AAS27440	cDNA encoding nove
34	104.6	4.4	7982	23	ABL02334	Drosophila melanog
35	102.8	4.4	3162	24	ABQ72637	Human MDR1T encodin
36	102.4	4.4	6411	23	ABL02196	Drosophila melanog
37	102.2	4.3	621	24	ABQ72602	Human MDR1T encodin
38	99.4	4.2	531	22	AAL22865	Human breast canc
39	98.4	4.2	2066	24	ABL60867	Human cytoskeleton
40	98	4.2	717	24	ABL90828	Human polynucleoti
41	95.8	4.1	659	22	AAS33845	Human colon cancer
42	90.2	3.8	544	22	AAH98023	Murine 7-transmemb
43	87	3.7	3260	23	ABL02197	Drosophila melanog
44	83.6	3.6	665	22	AAH07867	Human cDNA clone (
45	83.6	3.6	1296	24	ABK49898	Human cDNA encodin

ALIGNMENTS

RESULT 1	
ID	AAX79183 standard; DNA; 3442 BP.
XX	
AC	AAX79183;
XX	
DT	17-AUG-1999 (first entry)
XX	
DE	Human chondrocyte-derived gene CDEP.
XX	
KW	Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW	Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;
KW	drug; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9928458-A1.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-JP05348.
XX	
PR	27-NOV-1997; 97JP-0342060.
XX	
PA	(CHUS) CHUGAI SEIYAKU KK.
XX	
PI	Kato Y, Kawamoto T, Koyano Y;
XX	
DR	WPI; 1999-371117/31.
XX	
PT	P-PSDB; AAY07482.
Protein CDEP expressed in differentiated chondrocytes, and gene	

PT encoding it
XX PS Claim 5; Fig 1; 59pp; Japanese.
XX
CC This sequence represents the coding region for a protein (CDEP) expressed
CC in differentiated human foetal chondrocytes, which contains an ezrin-like
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.
CC The nucleic acid or protein can be used in the investigating and
CC treatment of cancers and arthritic diseases (including chronic rheumatoid
CC arthritis), or for screening of candidate anticancer drugs.
XX
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;
Query Match 100.0%; Score 2351; DB 20; Length 3442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 COTTGAGAAACCTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
DB 151 CTTTCAGGAAAACCTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210
QY 61 GAAATTCACAAAGAGCTCTCGGAGAGTGTCTGTGGATGAGTTGCAACCACTCAAC 120
DB 211 GAAATTCACAAAGAGCTCTCGGAGAGTGTCTGTGGATGAGTTGCAACCACTCAAC 270
QY 121 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 180
DB 271 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 330
QY 181 CTGGATCTCTAAACCACTTGTGAAACAGATTTAGAGGCCAAAGCAGCTGTGTTTAAAG 240
DB 331 CTGGATCTCTAAACCACTTGTGAAACAGATTTAGAGGCCAAAGCAGCTGTGTTTAAAG 390
QY 241 TTTGTGGTGAATCTTTCGCTCACCACACACAACTCCAGAGAACTCACAGGTAC 300
DB 391 TTTGTGGTGAATCTTTCGCTCACCACACAACTCCAGAGAACTCACAGGTAC 450
QY 301 CTGTTCCGCTGCAGGTGAAGCAGGACTTGGCTCAAGCAGGTTGACGTGTAATGACACC 360
DB 451 CTGTTCCGCTGCAGGTGAAGCAGGACTTGGCTCAAGCAGGTTGACGTGTAATGACACC 510
QY 361 AGCGAGCTCTCTGATTTACACATTTGTGCAATCTGAGATTTGGGATTTGATGAAGCC 420
DB 511 AGCGAGCTCTCTGATTTACACATTTGTGCAATCTGAGATTTGGGATTTGATGAAGCC 570
QY 421 TTGGACAGAGAGCACTTAGCAAAAATAAATACATACCTCAGCAAGACGCATAGAGAC 480
DB 571 TTGGACAGAGAGCACTTAGCAAAAATAAATACATACCTCAGCAAGACGCATAGAGAC 630
QY 481 AAAATCGTGGAAATTTACCATTAACACATTTGGACAAAACACAGCAATCAGATTTCCAG 540
DB 631 AAAATCGTGGAAATTTACCATTAACACATTTGGACAAAACACAGCAATCAGATTTCCAG 690
QY 541 CTCCTAGAGATTTGCCGTGATAGATGTATGGAATTCGCGTTGCACCGGCCCAAGAC 600
DB 691 CTCCTAGAGATTTGCCGTGATAGATGTATGGAATTCGCGTTGCACCGGCCCAAGAC 750
QY 601 AGGGAAGCAGCAAGATCAATCTGSCCGTTGCCAACACGGAATTTCTAGTTTTCAGGGT 660
DB 751 AGGGAAGCAGCAAGATCAATCTGSCCGTTGCCAACACGGAATTTCTAGTTTTCAGGGT 810
QY 661 TTCCTAAGATCAATGCTTCAACTTGGGCCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 720
DB 811 TTCCTAAGATCAATGCTTCAACTTGGGCCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 870
QY 721 CGCTTTCATCAAGCTCCGGCAGATGCCAATAGTGGTACCAGGATACCTTGGAAATTC 780
DB 871 CGCTTTCATCAAGCTCCGGCAGATGCCAATAGTGGTACCAGGATACCTTGGAAATTC 930
QY 781 CTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAAAATCTGTGTTGAACATCATGCC 840
DB 931 CTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGGAAAATCTGTGTTGAACATCATGCC 990

QY 841 TTCTTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCTCTCTTTAGCGGGG 900
DB 991 TTCCTTAGACTTTTGAAGAGCCCAAAAGCCCAAGCCCTCTCTTTAGCGGGG 1050
QY 901 TCATCATTTTCGGTTCAGTGGTTCGAGCTCAGAAAGAGGTTCTCGACTATGTTAAAGAGGA 960
DB 1051 TCATCATTTTCGGTTCAGTGGTTCGAGCTCAGAAAGAGGTTCTCGACTATGTTAAAGAGGA 1110
QY 961 GGACATAGAAGAGTGCAGTTTGAAGGAAGCAGCAAGAGATTCTATCCGAGGCCCT 1020
DB 1111 GGACATAGAAGAGTGCAGTTTGAAGGAAGCAGCAAGAGATTCTATCCGAGGCCCT 1170
QY 1021 GCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGAGCACCAGC 1080
DB 1171 GCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCTCAGCAGAGCACCAGC 1230
QY 1081 CTTTACATTTGGAGAAGGTGCGAATCTCCAGGGGCCAGAGCTGCCGGCGAGGAAGAA 1140
DB 1231 CTTTACATTTGGAGAAGGTGCGAATCTCCAGGGGCCAGAGCTGCCGGCGAGGAAGAA 1290
QY 1141 CCGAAGCTTTCCGCGGGAGCGGGGTGCGACCCGAGCCCTGCGCGAGGAGAGCCCC 1200
DB 1291 CCGAAGCTTTCCGCGGGAGCGGGGTGCGACCCGAGCCCTGCGCGAGGAGAGCCCC 1350
QY 1201 CCGGTAAACAGCAGCGGAGCGGCCCTCGGCGCCAGCGAGGAAGAGAGAGGTC 1260
DB 1351 CCGGTAAACAGCAGCGGAGCGGCCCTCGGCGCCAGCGAGGAAGAGAGAGGTC 1410
QY 1261 GTTAAAGATAGAGCCAGCAGAGTAACCTCAGCCCCCGCAGCCAAAGCAGAGGTCCCTG 1320
DB 1411 GTTAAAGATAGAGCCAGCAGAGTAACCTCAGCCCCCGCAGCCAAAGCAGAGGTCCCTG 1470
QY 1321 ACTGCGAGTCTCAGCTTTCGAGCTGTCTGTAACCTCGAGGGGGAGTGCCCTGCC 1380
DB 1471 ACTGCGAGTCTCAGCTTTCGAGCTGTCTGTAACCTCGAGGGGGAGTGCCCTGCC 1530
QY 1381 AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC 1440
DB 1531 AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCTCTCCCTTGATCAGC 1590
QY 1441 CCGTGTGTAATGACAGGCTTCCCGGGAGCGAGCATGAGGATGAGGGCGGAGGAAG 1500
DB 1591 CCGTGTGTAATGACAGGCTTCCCGGGAGCGAGCATGAGGATGAGGGCGGAGGAAG 1650
QY 1501 AGATTCCCACTGATAAGCGTACTTCTAGTAAGGAAGTGTCTACCAACGAGGGAACA 1560
DB 1651 AGATTCCCACTGATAAGCGTACTTCTAGTAAGGAAGTGTCTACCAACGAGGGAACA 1710
QY 1561 TATCTGAAGGATCTCGAAGTTATCATTCTCGTGGTTTCAGACACAGTGAAGCAAGAGAC 1620
DB 1711 TATCTGAAGGATCTCGAAGTTATCATTCTCGTGGTTTCAGACACAGTGAAGCAAGAGAC 1770
QY 1621 GCCATGCCGAAGCACTGAAAAGTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT 1680
DB 1771 GCCATGCCGAAGCACTGAAAAGTCTCATATTCGGAATTTTGAACCTTTGCACAAATTT 1830
QY 1681 CATCTAATTTTCTCAAGAAATTCAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAAT 1740
DB 1831 CATCTAATTTTCTCAAGAAATTCAGCAACGACTTGGCCCTGTGGGAAGCGCGCTCAAT 1890
QY 1741 GCCCAATACAGAGATTACCAAGAAATCGCGATGTCTGCTGAAGAACAATTCAGGGCATG 1800
DB 1891 GCCCAATACAGAGATTACCAAGAAATCGCGATGTCTGCTGAAGAACAATTCAGGGCATG 1950
QY 1801 AAGCACCTGGCGCTCAGCTGTGGAAGCAGACGAGGCCCTTGGAGGCCCTGGAGAAATGGA 1860
DB 1951 AAGCACCTGGCGCTCAGCTGTGGAAGCAGACGAGGCCCTTGGAGGCCCTGGAGAAATGGA 2010
QY 1861 ATCAAGAGCTCCCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGT 1920
DB 2011 ATCAAGAGCTCCCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGT 2070
QY 1921 TACCTACGCTCAACACCTTCTCCTGCGGCCACTGACCGGCTCATGCAAGCAG 1980

Db 2071 TACTACCGCTCAACACTTCTCCTCGCGCCACTGCACCGGCTCATGCACTACAAGCAG 2130
QY 1981 GTCTTGAGCGGTGTGCAAAACACCCCGCGAGCCACGCCGACTTCAGGGACTGCCGA 2040
Db 2131 GTCTTGAGCGGTGTGCAAAACACCCCGCGAGCCACGCCGACTTCAGGGACTGCCGA 2190
QY 2041 GCCGCTTTGGCAGAGATFCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATG 2100
Db 2191 GCCGCTTTGGCAGAGATFCAGGAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATG 2250
QY 2101 GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTGGATTGGCATTGACATCTTGTG 2160
Db 2251 GAGAAATTCAGAGCTGCACGAACTCAAGAAAGATTGGATTGGCATTGACATCTTGTG 2310
QY 2161 GTTCCGGGAAGGAGTTCTATCCCTCTGGGAGCTCAGCAAGCTCTCGGGGAAGGGGCTC 2220
Db 2311 GTTCCGGGAAGGAGTTCTATCCCTCTGGGAGCTCAGCAAGCTCTCGGGGAAGGGGCTC 2370
QY 2221 CAGCAGGCGATGTTCTTCTGTTCAACGAGCTCCTGCTATACAGAGCCGGGGCTGACG 2280
Db 2371 CAGCAGGCGATGTTCTTCTGTTCAACGAGCTCCTGCTATACAGAGCCGGGGCTGACG 2430
QY 2281 GCCTCAATCAGTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAG 2340
Db 2431 GCCTCAATCAGTTAAAGTCCAGGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAG 2490
QY 2341 AGCGAAGACGA 2351
Db 2491 AGCGAAGACGA 2501
RESULT 2
AAA08582
ID AAA08582 standard; DNA; 4687 BP.
XX
AC AAA08582;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.
XX
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 90..3227
FT /*tag= a
FT /product= CYSKP-2
XX
XX WO200017355-A2.
XX
XX 30-MAR-2000.
XX
XX 17-SEP-1999; 99WO-US21565.
XX
XX 18-SEP-1998; 98US-0172226.
XX 27-APR-1999; 99US-0131321.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
XX
XX WPI; 2000-283582/24.
DR P-PSDB; AAY91947.
XX
XX Human cytoskeleton associated proteins, used to treat cell
PT proliferative, autoimmune/inflammatory, vesicle trafficking,
PT neurological, cell motility, reproductive and muscle disorders

xx Claim 9; Page 101-102; 113pp; English.
xx
CC AAA08581-96 encode human cytoskeleton associated proteins 1 to 16
CC (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,
CC vesicle trafficking, neurological, cardiovascular, cell motility,
CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CYSKP (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
CC cancers, autoimmune/antifibrotic disorders such as allergies, anemia,
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CYSKP antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CYSKP
CC (claimed).
xx
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;
Query Match 99.9%; Score 2349.4; DB 21; Length 4687;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATCGTGATCACACCCAGGAGCATTT 60
Db 192 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATCGTGATCACACCCAGGAGCATTT 251
QY 61 GAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGACAGTTTGCACACCTCAAC 120
Db 252 GAAGTTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGACAGTTTGCACACCTCAAC 311
QY 121 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 180
Db 312 CTGCTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGG 371
QY 181 CTGGATCTCTAAACCCCATTTGAAACAGATTTAGAGGCCAAAGCACCGTTGTTGTTAAG 240
Db 372 CTGGATCTCTAAACCCCATTTGAAACAGATTTAGAGGCCAAAGCACCGTTGTTGTTAAG 431
QY 241 TTTGTCGTGAATTTCTTCGCCCTGACACACACACTCCAAGAAGACTCAAGGTAC 300
Db 432 TTTGTCGTGAATTTCTTCGCCCTGACACACACACTCCAAGAAGACTCAAGGTAC 491
QY 301 CTGTTCCGCTGTCAGGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACC 360
Db 492 CTGTTCCGCTGTCAGGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACC 551
QY 361 AGCGAGCTCTCTTGATTTTCACACATTTGCAATCTGAGATTGGGATTTTGATGAGCC 420
Db 552 AGCGAGCTCTCTTGATTTTCACACATTTGCAATCTGAGATTGGGATTTTGATGAGCC 611
QY 421 TTGCACAGAGACACTTAGCAAAAATAATATACATACCTCAGCAAGCAGCTAGAGGAC 480
Db 612 TTGCACAGAGACACTTAGCAAAAATAATATACATACCTCAGCAAGCAGCTAGAGGAC 671
QY 481 AAAATCGTGAATTTTCAACCATTAACCATTTGGACAAACACACAGCAATCAGATTTCCAG 540
Db 672 AAAATCGTGAATTTTCAACCATTAACCATTTGGACAAACACACAGCAATCAGATTTCCAG 731
QY 541 CTCCTAGAGATTGCCCGTCGGCTAGAGATGATGGAATCCGGTTGCCCGGCAAGGAC 600
Db 732 CTCCTAGAGATTGCCCGTCGGCTAGAGATGATGGAATCCGGTTGCCCGGCAAGGAC 791
QY 601 AGGGAAGGCACGAGATCAATCTGGCCGTTGCCAAGCAGGATTTCTAGTTTTCAGGCT 660
Db 792 AGGGAAGGCACGAGATCAATCTGGCCGTTGCCAAGCAGGATTTCTAGTTTTCAGGCT 851
QY 661 TTCACCTAAGATCAATGCCTTCAACTGGGCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 720
Db 852 TTCACCTAAGATCAATGCCTTCAACTGGGCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAG 911
QY 721 CGCTTTCTCATCAAGCTCCGGGCGAGATGCCAATAGTGCCTACAGGATACCTTTGGAATTC 780

Db	912	CGCTTTTTCATCAAGCTTCGGCCAGATGCCAATAGTCGTACCAAGGATACCTTGGAAATTC	971
Qy	781	CTGATGCCAGTCGGGATTTCTGCAAGTCCTTCTTGGAATAATCTGTGTGAACATCATGCC	840
Db	972	CTGATGCCAGTCGGGATTTCTGCAAGTCCTTCTTGGAATAATCTGTGTGAACATCATGCC	1031
Qy	841	TTCTTTTAGACTTTTGAAGAGCCCAAAACAAAGCCCAAGCCGCTCTCTTTAGCCGGGG	900
Db	1032	TTCTTTTAGACTTTTGAAGAGCCCAAAACAAAGCCCAAGCCGCTCTCTTTAGCCGGGG	1091
Qy	901	TCATCATTTCCGTTTCAGTTCGGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGA	960
Db	1092	TCATCATTTCCGTTTCAGTTCGGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGA	1151
Qy	961	GGACATAAGAAGTCGAGTTTGAAGAGGAAGCACACCAAGATTCTATCTCCGAGGCGCTT	1020
Db	1152	GGACATAAGAAGTCGAGTTTGAAGAGGAAGCACACCAAGATTCTATCTCCGAGGCGCTT	1211
Qy	1021	GCCTACAGCCTACAGAACTGAATTCGGAAGTGCTGGAGCAGTCTCAGCAGAGCACCAGC	1080
Db	1212	GCCTACAGCCTACAGAACTGAATTCGGAAGTGCTGGAGCAGTCTCAGCAGAGCACCAGC	1271
Qy	1081	CTTACATTTCCGAAAGTCCGGAATCTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGAA	1140
Db	1272	CTTACATTTCCGAAAGTCCGGAATCTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGAA	1331
Qy	1141	CCGAAGGTTTCCGCCGGGAGCCGGGTCGCACCCAGCCCTCGCCGCGAGGAGAAGCCCC	1200
Db	1332	CCGAAGGTTTCCGCCGGGAGCCGGGTCGCACCCAGCCCTCGCCGCGAGGAGAAGCCCC	1391
Qy	1201	CGGGTTAACAGCAGCGGACGGAGCCGCTCGGGCCCGACCGAGGAGAGGAGGAGTCC	1260
Db	1392	CGGGTTAACAGCAGCGGACGGAGCCGCTCGGGCCCGACCGAGGAGAGGAGGAGTCC	1451
Qy	1261	GTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCAGCCACAGCAGAGGCTCCCTG	1320
Db	1452	GTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCAGCCACAGCAGAGGCTCCCTG	1511
Qy	1321	ACTGGCAGTCCTCACCCTTTCGAGGTGTCGTGTGAACCTCGCAGGGGGAGTGGCCCTGCC	1380
Db	1512	ACTGGCAGTCCTCACCCTTTCGAGGTGTCGTGTGAACCTCGCAGGGGGAGTGGCCCTGCC	1571
Qy	1381	AACGTGACCTTGTCTCCCAACCTGAGCCCGGACACCAAGCAGGCTCTCCCTTGATCAGC	1440
Db	1572	AACGTGACCTTGTCTCCCAACCTGAGCCCGGACACCAAGCAGGCTCTCCCTTGATCAGC	1631
Qy	1441	CCGCTGCTGAATCACCAGGCTTCCCGCGGAGGAGATGAGGATGAGGCGCGGAGGAAG	1500
Db	1632	CCGCTGCTGAATCACCAGGCTTCCCGCGGAGGAGATGAGGATGAGGCGCGGAGGAAG	1691
Qy	1501	AGATTCCCACTCATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCGCAGCGAACA	1560
Db	1692	AGATTCCCACTCATAAAGCGTACTTCATAGCTAAGGAAGTGTCTACACCGCAGCGAACA	1751
Qy	1561	TATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGTGACGAAGAGGAC	1620
Db	1752	TATCTGAAGGATCTCGAAGTTATCACTTCGTGGTTTCAGAGCAGTGACGAAGAGGAC	1811
Qy	1621	GCCATGCCGAAGACTGAAAAGTCTCATATTTCCGAAATTTTGAACCTTTGCACAAATTT	1680
Db	1812	GCCATGCCGAAGACTGAAAAGTCTCATATTTCCGAAATTTTGAACCTTTGCACAAATTT	1871
Qy	1681	CATACATAATTTCTCAGGAATTCAGCAACGACTTGCCTGTGGGAAGCCGCTCAAT	1740
Db	1872	CATACATAATTTCTCAGGAATTCAGCAACGACTTGCCTGTGGGAAGCCGCTCAAT	1931
Qy	1741	GCCCAATCAGATTAACCAAGAAATCGCGGATGTCATGCTGAAGAACAATTCAGGGCATG	1800
Db	1932	GCCCAATCAGATTAACCAAGAAATCGCGGATGTCATGCTGAAGAACAATTCAGGGCATG	1991
Qy	1801	AAGCACTGCGGCTCACCTGTGGGAAGCACAGCAGGCGCTTGGAGGCGCTTGAGAATTGA	1860

Db	1992	AAGCACC	TGCGGGCTCACCTGTGGAA	GACACACG	GAGGCC	TTGGAGGCC	CTGGAGAA	TGGA	2051
Qy	1861	ATCAAGAG	CTCCGCGGGCTGGAGAA	CTTCTGCAGAG	ACTTTTGAGCTG	CACAGAA	GCTGTGT		1920
Db	2052	ATCAAGAG	CTCCGCGGGCTGGAGAA	CTTCTGCAGAG	ACTTTTGAGCTG	CACAGAA	GCTGTGT		2111
Qy	1921	TACCTAG	CGCTCAACACCTTCTCT	TGGGGCCACTGCAC	CGGCTCATGCACT	ACAAGCA	G		1980
Db	2112	TACCTAG	CGCTCAACACCTTCTCT	TGGGGCCACTGCAC	CGGCTCATGCACT	ACAAGCA	G		2171
Qy	1981	GTCTTG	AGCGGCTGTGC	AAACACACCG	CGCGAGCGG	CACTTCAGG	CACTGCCGA		2040
Db	2172	GTCTTG	AGCGGCTGTGC	AAACACACCG	CGCGAGCGG	CACTTCAGG	CACTGCCGA		2231
Qy	2041	GC	CGCTTTGGCAGAGAT	CACGAGAGTGTGG	CACAGCTCCAC	GGTACGAT	CAATCAAGAT		2100
Db	2232	GCC	CTTTGGCAGAGAT	CACGAGAGTGTGG	CACAGCTCCAC	GGTACGAT	CAATCAAGAT		2291
Qy	2101	GAGA	ATTTCCAGAAAGCTGC	ACGAACCTCAAGA	AAAGATTGATTGG	CAATTTG	GTG		2160
Db	2292	GAGA	ATTTCCAGAAAGCTGC	ACGAACCTCAAGA	AAAGATTGATTGG	CAATTTG	GTG		2351
Qy	2161	GT	TCCGGGAAGGAGTTCAT	CTCGTGTGGCAG	CGCTCAGCAAG	CTCTCGGG	AAGGGGCTC		2220
Db	2352	GT	TCCGGGAAGGAGTTCAT	CTCGTGTGGCAG	CGCTCAGCAAG	CTCTCGGG	AAGGGGCTC		2411
Qy	2221	CAG	CAGCGCATGTTCTTCT	CTGTTCACGAG	CTCTGCTATACAG	AGCGCGGGGCTC	ACG		2280
Db	2412	CAG	CAGCGCATGTTCTTCT	CTGTTCACGAG	CTCTGCTATACAG	AGCGCGGGGCTC	ACG		2471
Qy	2281	GC	CTCAATCAGTTTAAAG	TCCACGGG	CAGCTCCCGCTCTAT	TGGCATGAC	GATTTAGGAG		2340
Db	2472	GC	CTCAATCAGTTTAAAG	TCCACGGG	CAGCTCCCGCTCTAT	TGGCATGAC	GATTTAGGAG		2531
Qy	2341	AG	CGAAGACGA						2351
Db	2532	AG	CGAAGACGA						2542

RESULT 3	
AAS64760	
ID	AAS64760 standard; cDNA; 3187 BP.
XX	XX
AC	AAS64760;
XX	XX
DT	13-FEB-2002 (first entry)
XX	XX
DE	DNA encoding novel human diagnostic protein #564.
XX	XX
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200175067-A2.
XX	XX
PD	11-OCT-2001.
XX	XX
PF	30-MAR-2001; 2001WO-US08631.
XX	XX
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	XX
PA	(HYSE-) HYSEQ INC.
XX	XX
PI	Drganac RT, Liu C, Tang YT;
XX	XX
DR	WPI: 2001-639362/73.
DR	P-PSDB; ABG00573.
XX	XX
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess

PT biodiversity -
XX Claim 1; SEQ ID No 564; 103pp; English.
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3187 BP; 810 A; 874 C; 876 G; 627 T; 0 other;

Query Watch 99.3%; Score 2335.2; DB 23; Length 3187;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2348; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 CCTCTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGGATGACACACCCAGGAGCATTT 60
DB 151 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATCGTGGATGACACACCCAGGAGCATTT 210

QY 61 GAAGTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAATTTGCAACCACTCAAC 120
DB 211 GAAGTCCACAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAATTTGCAACCACTCAAC 270

QY 121 CTCGTGGAAGGTGACATTTTGGCCCTCGAGTTTCCTGATGATCACAAAAGATCAGCGTGTGG 180
DB 271 CTCGTGGAAGGTGACATTTTGGCCCTCGAGTTTCCTGATGATCACAAAAGATCAGCGTGTGG 330

QY 181 CTGGATCTCCTTAAACCCATTTGTGAACAGATTAGAAGGCCAAAGCACCGTTGTTGTTAAG 240
DB 331 CTGGATCTCCTTAAACCCATTTGTGAACAGATTAGAAGGCCAAAGCACCGTTGTTGTTAAG 390

QY 241 TTTGTGGTGAATTTCTTCCGCTGACCAACACAACTCCAAGAAGAACTCAAAAGGTAC 300
DB 391 TTTGTGGTGAATTTCTTCCGCTGACCAACACAACTCCAAGAAGAACTCAAAAGGTAC 450

QY 301 CTGTTCCGCTCGAGTGAAGCAGGACTTGGCTCAAGCAGGTGACGTGTAATGACACC 360
DB 451 CTGTTCCGCTCGAGTGAAGCAGGACTTGGCTCAAGCAGGTGACGTGTAATGACACC 510

QY 361 AGCGCAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGGATTTTGATGAAGCC 420
DB 511 AGCGCAGCTCTCTTGATTTTACACATTTGCAATCTGAGATTGGGGATTTTGATGAAGCC 570

QY 421 TTGGCAGAGAGACATTTAGCAAAAAATAATACATACCTCAGCAAGACGCACTTAGAGGAC 480
DB 571 TTGGCAGAGAGACATTTAGCAAAAAATAATACATACCTCAGCAAGACGCACTTAGAGGAC 630

QY 481 AAAATCGTGAATTTTCAACCATRACCATTTGGCAACACACCAGCAGCAATTCAGTTCCAG 540
DB 631 AAAATCGTGAATTTTCAACCATRACCATTTGGCAACACACCAGCAGCAATTCAGTTCCAG 690

QY 541 CTCCTAGAGATTGCCCGCTCGGTAGAGATTGATGAATCCGGTTGCACCCGGCCCAAGGAC 600
DB 691 CTCCTAGAGATTGCCCGCTCGGTAGAGATTGATGAATCCGGTTGCACCCGGCCCAAGGAC 750

QY 601 AGGGAAGGCAGGAAGATCAATCTGGCCGTGGCCAAACACGGGAATTTCTAGTGTTCAGGGT 660

DB 751 AGGGAAGGCAGGAAGATCAATCTGGCCGTTGCCAACACGGGAATTTCTAGTGTTCAGGGT 810
QY 661 TTCACATAAGATCAATGCCCTTCAACTGGGCCAAGGTCCGGAAGCTTGACGCTTCAAGAGAAG 720
DB 811 TTCACATAAGATCAATGCCCTTCAACTGGGCCAAGGTCCGGAAGCTTGACGCTTCAAGAGAAG 870
QY 721 CGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGCCTTACCAGGATACCTTTGGAATTC 780
DB 871 CGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGTGCCTTACCAGGATACCTTTGGAATTC 930
QY 781 CTGATGGCAGTCGGGATTTCTGCAAGTCTCTTCTGAAAATCTGTGTGAACATCATGCC 840
DB 931 CTGATGGCAGTCGGGATTTCTGCAAGTCTCTTCTGAAAATCTGTGTGAACATCATGCC 990
QY 841 TTCCTTTAGACTTTTTCAGAGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCCGGGG 900
DB 991 TTCCTTTAGACTTTTTCAGAGCCCAACCAAGCCCAAGCCCGTCTCTTTAGCCCGGGG 1050
QY 901 TCATCATTTTCGGTTCAGTGTGCGACTCAGAAGCAGGTTCTCGACTATGTAAAGAAGGA 960
DB 1051 TCATCATTTTCGGTTCAGTGTGCGACTCAGAAGCAGGTTCTCGACTATGTAAAGAAGGA 1110
QY 961 GGACATAAGAAGTGCAGTTTGAAGGAAGCACAGCAAGATTCTATPCGGAGCCCTT 1020
DB 1111 GGACATAAGAAGTGCAGTTTGAAGGAAGCACAGCAAGATTCTATPCGGAGCCCTT 1170
QY 1021 GCTTCACAGCTTACAGAACTGAATTCGGAAGTCTGAGCAGTCTCAGCAGACACACAGC 1080
DB 1171 GCTTCACAGCTTACAGAACTGAATTCGGAAGTCTGAGCAGTCTCAGCAGACACACAGC 1230
QY 1081 CTTTACATTTGGAGA-AGGTGCCGAATCTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGA 1139
DB 1231 CTTTACATTTGGAGAGAGTGCCTGAATTTCCAGGGGCCAGAGCTGCCGGCAGGAAAGGA 1290
QY 1140 ACCGAAGTTCCTCGCCGGGAGCCGGGTGCGACCCAGCCCTGCGCCCGAGGAGAAGCC 1199
DB 1291 ACCGAAGTTCCTCGCCGGGAGCCGGGTGCGACCCAGCCCTGCGCCCGAGGAGAAGCC 1350
QY 1200 CCGGGTAAACAGCAGCGGACGGAGCCGCTCGGCCCCAGCCAGCAGCAGGAGGAGGT 1259
DB 1351 CCGGGTAAACAGCAGCGGAGCGGAGCCGCTCGGAGCCCGCAGGAGGAGGAGGT 1410
QY 1260 CGTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCGCCAGCAGCAGCAGGCTCCCT 1319
DB 1411 CGTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCTCGCAGCCAGCAGCAGGCTCCCT 1470
QY 1320 GACTGCGAGTCTCAGCTTTTCCGAGCTGCTGTGAATCTCGCAGGGGGAGTGGCCCTGC 1379
DB 1471 GACTGCGAGTCTCAGCTTTTCCGAGCTGCTGTGAATCTCGCAGGGGGAGTGGCCCTGC 1530
QY 1380 CAACGTGACCTTGTCTCCCAACCTGAGCCCCGACCAAGCAGCCCTCTCCCTTGATCAG 1439
DB 1531 CAACGTGACCTTGTCTCCCAACCTGAGCCCCGACCAAGCAGCCCTCTCCCTTGATCAG 1590
QY 1440 CCGCTGCTGAATGACACAGCCCTGCCCCGGAGCAGCATGAGGATGAGGGCCGAGGAA 1499
DB 1591 CCGCTGCTGAATGACACAGCCCTGCCCCGGAGCGAGATGAGGATGAGGGCCGAGGAA 1650
QY 1500 GAGATTCCCAACTGATAAAGCGTACTTTCATAGTAAGGAGTGTCTTACCACCGAGGCAAC 1559
DB 1651 GAGATTCCCAACTGATAAAGCGTACTTTCATAGTAAGGAGTGTCTTACCACCGAGGCAAC 1710
QY 1560 ATATCTGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGACACAGTGAAGCAAGGA 1619
DB 1711 ATATCTGAAGGATCTCGAAGTATATCACTTCGTGGTTTCAGACACAGTGAAGCAAGGA 1770
QY 1620 CGCCATGCCGGAAGCAGTCAAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAAT 1679
DB 1771 CGCCATGCCGGAAGCAGTCAAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAAT 1830
QY 1680 TCATACTAATTTTCTCAAGGAATTCAGCAACGACTTGGCCCTGTGGGAAGGCGCTCAAA 1739

Db 1831 TCATACTAATTTCTCAAGAAATTTGAGCAACGACTTGGCCCTGTGGAGGCGGCTCAAA 1890
QY TGCCCAAATCAGAGATTACCAAGAATCGCGATGTGCTGCTGAAGAACATTCAGGGCAT 1799
Db 1891 TGCCCAAATCAGAGATTACCAAGAATCGCGATGTGCTGCTGAAGAACATTCAGGGCAT 1950
QY 1800 GAAGCACCTTGGGGCTCACCTGTGTGAAGCACACGCGAGGCTTGGAGGCGCTGGAGAATGG 1859
Db 1951 GAAGCACCTTGGGGCTCACCTGTGTGAAGCACACGCGAGGCTTGGAGGCGCTGGAGAATGG 2010
QY 1860 AATCAGAGCTCCCGGGCTGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTG 1919
Db 2011 AATCAGAGCTCCCGGGCTGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTG 2070
QY 1920 TTACCTACCGCTCAACACCTTCTCTCTCGGCGCACTGCACCGGCTCATGCATACAAACA 1979
Db 2071 TTACCTACCGCTCAACACCTTCTCTCTCGGCGCACTGCACCGGCTCATGCATACAAACA 2130
QY 1980 GGTCTTGGAGCGGCTGTGCAAAACACACCGCGCGAGCCACGCGGACTTCAGGGACTGCGG 2039
Db 2131 GGTCTTGGAGCGGCTGTGCAAAACACACCGCGCGAGCCACGCGGACTTCAGGGACTGCGG 2190
QY 2040 AGCCGCTTTGGCAGAGATCAGGAGATGTGTGCAGAGCTTCCAGGTTACATGATCAAGAT 2099
Db 2191 AGCCGCTTTGGCAGAGATCAGGAGATGTGTGCAGAGCTTCCAGGTTACATGATCAAGAT 2250
QY 2100 GGAGAAATTTCCAGAGCTGCAGAACTCAGAAAGATTTGATTGGCATTGACAAATCTTCT 2159
Db 2251 GGAGAAATTTCCAGAGCTGCAGAACTCAGAAAGATTTGATTGGCATTGACAAATCTTCT 2310
QY 2160 GGTTCGGGAAGGAGTTTCATCCGTTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCT 2219
Db 2311 GGTTCGGGAAGGAGTTTCATCCGTTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCT 2370
QY 2220 CCAGCAGCGCATGTTCTTCTGTTCAACACGCTCTGCTATACAGAGCCGGGGCTGAC 2279
Db 2371 CCAGCAGCGCATGTTCTTCTGTTCAACACGCTCTGCTATACAGAGCCGGGGCTGAC 2430
QY 2280 GGCTCCAACTCAGTTTAAAGTCCACGGGAGCTCCGCGCTATGTCATGACGATTGAGCA 2339
Db 2431 GGCTCCAACTCAGTTTAAAGTCCACGGGAGCTCCGCGCTATGTCATGACGATTGAGCA 2490
QY 2340 GAGCGAAGACGA 2351
Db 2491 GAGCGAAGACGA 2502
RESULT 4
AAC98992
ID AAC98992 standard; cdNA; 3094 BP.
XX
AC AAC98992;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO20005320-A1.
XX
XX 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX

PR 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Ruben SM;
PI WPI; 2000-579444/54.
DR P-PSDB; AAB54227.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX Claim 1; Page 664-665; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;
Query Match 32.4%; Score 762; DB 21; Length 3094;
Best Local Similarity 100.0%; Pred. No. 3.2e-205;
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1590 GTGGTTTCAGACAGTGAGCAAGAGAGACGCGCATGCCGGAAGCTCAAAAGTCTCAT 1649
Db 62 GTGGTTTCAGACAGTGAGCAAGAGAGACGCGCATGCCGGAAGCTCAAAAGTCTCAT 121
QY 1650 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATGAGCA 1709
Db 122 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAAATGAGCA 181
QY 1710 ACGACTTGCCCTGTGGAGGCGGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 1769
Db 182 ACGACTTGCCCTGTGGAGGCGGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 241
QY 1770 CGATGTCTGCTGAAGAACATTCAGGCGATGAAGCACCTGGCGGCTCACTGTGGAAGCA 1829
Db 242 CGATGTCTGCTGAAGAACATTCAGGCGATGAAGCACCTGGCGGCTCACTGTGGAAGCA 301
QY 1830 CAGCGAGGCTTTGGAGGCGGCTGGAGAAATGGAATCAAGAGCTCCCGGCGGCTGGAGAACTT 1889
Db 302 CAGCGAGGCTTTGGAGGCGGCTGGAGAAATGGAATCAAGAGCTCCCGGCGGCTGGAGAACTT 361
QY 1890 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 1949
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGGG 421
QY 1950 GCCACTGCACCGGCTCATGCTACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACCCACC 2009
Db 422 GCCACTGCACCGGCTCATGCTACTACAAGCAGGTCTCTGGAGCGGCTGTGCAACACCCACC 481
QY 2010 GCCGAGCCACGCCGACTTCAGGAGCTTGCAGGCGGCTTTTGGCAGAGATCACGAGATGTT 2069

|||||
Db 482 GCCGAGCCACGCGCTTACGAGCTGCGGAGCCCTTTGGCAGAGATCACGGATGCT 541
QY 2070 GGCACAGCTCCACGGTACGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA 2129
Db 542 GGCACAGCTCCACGGTACGATCAAGATGAGAAATTTCCAGAAGCTGCAGAACTCAA 601
QY 2130 GAAAGATTGATTGGCATTGACAACTTTGTGTTCCGGGAAGGGAGTTTCATCCGCTGGG 2189
Db 602 GAAAGATTGATTGGCATTGACAACTTTGTGTTCCGGGAAGGGAGTTTCATCCGCTGGG 661
QY 2190 CAGCCTCAGCAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTCTTTCAACGA 2249
Db 662 CAGCCTCAGCAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTCTTTCAACGA 721
QY 2250 CGTCTGCTATACAGCAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGCA 2309
Db 722 CGTCTGCTATACAGCAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGCA 781
QY 2310 GTCCTCGCTATGATGATGACGATTGAGGAGAGCGAAGACGA 2351
Db 782 GTCCTCGCTATGATGATGACGATTGAGGAGAGCGAAGACGA 823
RESULT 5
AAS64761/c
ID AAS64761 standard; cDNA; 1824 BP.
XX AC AAS64761;
XX
XX 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #565.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG00574.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 565; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1824 BP; 409 A; 477 C; 486 G; 452 T; 0 other;
Query Match 10.8%; Score 254; DB 23; Length 1824;
Best Local Similarity 100.0%; Pred. No. 3.5e-61;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1060 CAGTCTCAGCAGAGCAGCAGCCTTACATTTGGAGAAGGTTCCTCCAGGGGGCCAG 1119
Db 1623 CAGTCTCAGCAGAGCAGCAGCCTTACATTTGGAGAAGGTTCCTCCAGGGGGCCAG 1564
QY 1120 AGCTGCCGCGAGGAAAGAACCCGAGGTTTCCTCCGGGGAGCGGGGTTCGCACCCGAGC 1179
Db 1563 AGCTGCCGCGAGGAAAGAACCCGAGGTTTCCTCCGGGGAGCGGGGTTCGCACCCGAGC 1504
QY 1180 CCTGCGCGAGGAGAGCCCGCGGGTTAACAGCAGCGAGCGCGCCCTTCGGCGCC 1239
Db 1503 CCTGCGCGAGGAGAGCCCGCGGGTTAACAGCAGCGAGCGCGCCCTTCGGCGCC 1444
QY 1240 ACGGAGGAAGAGGAGGAGTCTTTAAGGATAGGAGCCAGCAGAGTAAACCTCAGCCCCCG 1299
Db 1443 ACGGAGGAAGAGGAGGAGTCTTTAAGGATAGGAGCCAGCAGAGTAAACCTCAGCCCCCG 1384
QY 1300 CAGCCAAGCAGAGG 1313
Db 1383 CAGCCAAGCAGAGG 1370
RESULT 6
ABL02335
ID ABL02335 standard; cDNA; 1220 BP.
XX
XX AC ABL02335;
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1487.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR P-PSDB; ABB58232.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX

PS Claim 1; SEQ ID NO 1487; 21pp + Sequence Listing; English.

xx The invention relates to an isolated nucleic acid detection reagent

cc capable of detecting 1000 or more genes from Drosophila. The invention is

cc useful in developmental biology and in elucidating cell signalling and

cc cell-cell interactions in higher eukaryotes for the development of

cc insecticides, therapeutics and pharmaceutical drugs. The invention

cc discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

cc sequences (ABL01840-ABL16175) and the encoded proteins

cc (ABB57737-ABB72072).

cc The sequence data for this patent did not form part of the printed

cc specification, but was obtained in electronic format directly from WIPO

cc at ftp.wipo.int/pub/published_pct_sequences.

xx

xx Sequence 1220 BP; 309 A; 327 C; 329 G; 255 T; 0 other;

Query Match 9.78; Score 228; DB 23; Length 1220;

Best Local Similarity 55.18; Pred. No. 6.6e-54;

Matches 557; Conservative 0; Mismatches 435; Indels 18; Gaps 5;

7 GGAATACTCGTTCATCAAAATCCAGATCGTGGATGACACCCAGGAGGCATTTTGAAGTT 66

198 GGCATAAAGCTGACTGTTTCGATCCAGATGCTGGACGACTCGATCACCATTGTTCCAAGTA 257

67 CCACAAAGAGTCTCTGGGAAGTGTCTGATGACGTTTCCAAACCACTCAACCTCGTG 126

258 CAGGCTAAAGCACTGGGCGCGTGTGTTGAACAGGATGCGCGTCAATTTACTG 317

127 GAAGTGACTATTTGGCGTCGAGTTTCTCTGATCA---CAAAAGATCACCAGTGGCGTG 183

318 GAGGCGGACTACTTGGGCTTGAGTACCAGGAAGTCTCCACATACCAATATCTGGCTG 377

184 GATCTCTTAAACCCATTTGTGAACACAGATTAGAAGGCGCAAGCGTGTGTTTAAG--- 240

378 GATCTGGAGAGCAATGAACCGCCAGTGGGCGCTCTCTTATCGATCCAGTCTGGC 437

241 TTTGGTGAATTTCTTCCGCTGACACACACAACTCCAAAGAACTCACAAGGTAC 300

438 TTCTGCATCAAGTTCTATACACCGGATCCAGCGCAATTTGGAAGAGGAGTACACCGAT 497

301 CTGTTGCGCTGAGGTTGAAGCAAGCACTTGCTCAAGCGAGTTCAGTGTGAATGACACC 360

498 TTGTTCTCTGCAATCAACAGAGACTGGCCACAGGACGCTGCGAGTGACAGATTAAC 557

361 AGCCGAGCTCTTGTATTCACATTTGCAATTCGAGATTGGGATTTTGTAT---GAAG 418

558 ACGGCGGCTTAATGGCAAGTACATTGTACAGGCGTATCGGCGGACTTCTTCCGGAG 617

419 CTTGGACAGAGAGCTTTAGCAAAATAAATACATACCT-----CAGCAAGACGCA 471

618 GATTATCCCGACCATCTATCTGTCTCTGATGCTTTGTGCCCAACCCAGGAGGCCACC 677

472 CTAGAGGACAAATCGTGGAAATTCACCATACACATGTCAGCAACACACAGCAGCAATCA 531

678 ATGACGGAAGATTTATGAGAACACCAAAAGGATGTTGGTCAATCCCGCGGAGGCG 737

532 GATTTCACAGCTAGAGATTGCCGCTCGGCTAGAGATGTATGGAATCCGGTTGACCCCG 591

738 GACCTTAACTTTGGACACGCTCGGAGGTGTGAGCTGTATGCGATGAAATGCATCCG 797

592 GCCAAGGACAGGAGGACCAAGATCAATCTGGCCGTTGGCCACACACGGGAATTTAGTG 651

798 GCGAAGGATGTGGAAGGGGTGCCGCTTAATCTGGGTGTGGCCCATCATGGCATCACAGTC 857

652 TTTACAGGTTTCACTTAAGATCAATGCTTCAACTTGGGCGCAAGGTGCGGAAGCTGAGCTTC 711

858 TTCAGAACATACCGGATCAACACCTTCTGTTGGGTAGATACGGAATTTCTTTC 917

712 AAGAGAACCGCTTCTCATCAAGTCCGCGCAGATAGTGCATAGTGCATACAGGATACC 771

918 AAGCGCAAGCATCTCTGTTCAACTGATCGGAGGATATGGATATACAGGATACC 977

772 TTGGAATTCGTAGTGGCCAGTCTGGGATTTCTGCAAGTCTTCTTGAAAAATCTGTGTGAA 831

Db 978 GTGAGTCTTCTTTCGAGGGTCGCAACGAGTGCAGAACTTCTGAAAAATGCGTGCA 1037

Qy 832 CATCATCGCTCTTTAGACTTTTGAAGAGCCAA---ACCAAGCCCAAGCCGCTGCTC 888

Db 1038 AATCAGGATTTCTTCGATGACGCTGCGGTACAAATACGCCAGCGCAAACTCGCGTT 1097

Qy 889 TTTAGCCGGGGTTCATCATTTTCGTTTCAGTGTGCGACTCAGAACGAGTTCCTCGACTAT 948

Db 1098 CTCTCGCGGGTAGTTTCATTCCTCGCTATAGCGGAAAAACCCAGAGCAGATTATCGAGTTC 1157

Qy 949 GTTAAAGAGGAGGACATAGAAGGTGCGAGTTTGAAGGAAGCAGCAGCA 998

Db 1158 GTTCGCAAAATTTATGTGAAGCGCCAAAACCTTCCAAGGTATATAAACGAA 1207

RESULT 7

AAA42150/c

ID AAA42150 standard; cDNA; 241 BP.

XX

AC AAA42150;

DT 21-AUG-2000 (first entry)

XX

Human secreted expressed sequence tag SEQ ID NO:890.

Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;

expressed sequence tag; EST; probe; chemotactic; proliferative;

immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;

antiviral; osteopathic; antiasthmatic; vulnery; antiparkinsonian;

antiulcer; antiepileptic; neuroprotective; nootropic; antipsoriatic;

cerebroprotective; anticonvulsant; antidepressant; gene therapy;

vaccine; autoimmune disorder; multiple sclerosis; allergic condition;

insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

central nervous system disorder; Alzheimer's disease; stroke;

Parkinson's disease; Huntington's disease; coagulation disorder;

haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

tumour; infection; depression; psoriasis; ss.

XX

Homo sapiens.

OS

XX

WO200021990-A1.

PN

XX

20-APR-2000.

PD

XX

15-OCT-1999; 99WO-US24205.

PF

XX

15-OCT-1998; 98US-0104435.

PR

XX

(GEMY) GENETICS INST INC.

PA

XX

Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M;

PT

XX

WPI; 2000-317937/27.

DR

XX

Isolated polynucleotides, and encoded proteins, comprising secreted

expressed sequence tags (sESTs), useful for treating various disorders

such as autoimmune, infectious, and central nervous system disorders -

Claim 1; Page 354; 618pp; English.

PS

XX

AAAA1261 to AAA43419 represent specifically claimed secreted expressed

sequence tags (sESTs), isolated from human, mouse, xenopus and rat

tissue sources. The sESTs can have a range of activities depending on

the tissues they were isolated from. The activities include:

CC chemotactic; proliferative; immunomodulatory; haematopoietic;

CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;

CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;

CC antiasthmatic; vulnery; antileuk; osteopathic; neuroprotective;

CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;

anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA43420 to AAA43425 represent linker variants which are given in the exemplification of the present invention.

Sequence 241 BP; 59 A; 58 C; 58 G; 66 T; 0 other; XX
SQ

Query Match 9.5%; Score 223.6; DB 21; Length 241;
Best Local Similarity 98.3%; Pred. NO. 4.6e-53;
Matches 226; Conservative 0; Mismatches 4; Indels 0;

Qy	145	CTCAGATTCTGATCAGAAAAAGATCACGGTGTGGCTGGATCTCTAAAAACCCATTGTG	204
Db	241	CTCAGATTCTGATCAGAAAAAGATCACGGTGTGGCTGGATCTCTAAAAACCCATTGTG	182
Qy	205	AAACAGATTAGAGGCCAAGACAGTTGCTTAAATTGTGGTGAATTTCTTCCGGCT	264
Db	181	AAACAGATTAGAGGCCAAGACAGTTGCTTAAATTGTGGTGAATTTCTTCCGGCT	122
Qy	265	GACCACACACAACCTCCAGAGAACTCACAGGTACCTGTTCCGGCTGCAGGTGAAGCAG	324
Db	121	GACCACACACAACCTCCAGAGAACTCACAGGTACCTGTTCCGGCTGCAGGTGAAGCAG	62
Qy	325	GACTTGGCTCAAGCAGGTTGACGTGTAATGACACACGCGCAGCTTCTCTT	374
Db	61	GACTTGGCTCAAGCAGGTTGACGTGTAATGACACACGCTTTAGGCCTCTT	12

RESULT 8

ABV30214
ID ABV30214 standard; cDNA; 4922 BP.
XX
AC ABV30214;

DT 16-SEP-2002 (first entry)

Human prostate expression marker cDNA 30205.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

KW pharmacogenomic marker; gene; ss:

Homo sapiens.

XX
PN WO200160860-A2.

23-AUG-2001.

AA
PF 20-FEB-2001: 2001WO-US05171.

17-FEB-2000; 2000US-183319P.

PR	16-MAR-2000; 2000US-189862P.
PR	25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000S-255281P.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE:

WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of

prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 6547-6548; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

- (a) assessing whether a patient is afflicted with prostate cancer;
- (b) monitoring the progression of prostate cancer in a patient;
- (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
- (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
- (e) selecting a composition for inhibiting prostate cancer in a patient;
- (f) assessing the prostate cell carcinogenic potential of a compound;
- (g) determining whether prostate cancer has metastasized in a patient;
- (h) assessing the aggressiveness or indolence of prostate cancer in a patient;
- (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX
SQ Sequence 4922 BP; 1406 A; 1092 C; 1164 G; 1258 T; 2 other;

Query Match 8.2%; Score 192.4; DB 23; Length 4922;
Best Local Similarity 52.8%; Pred. No. 1.9e-43;
Matches 490; Conservative 0; Mismatches 426; Indels 12;

15 CGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGGCATTGGAAGTTCCACAAAG 74

Db 798 CATGCACTGCAAGGTTTCTTTGTTGGATGCACACAGTTTATGAATGTGTTGTGGAGAAACA 857

QY 75 AGCTCCTGGGAAGGTGCTGCTGGATGCAGTTTGAACCACCTCAACCTCGTGGGAAGGTGA 134

Db 858 TGCTAAGGGACAAGATTTCGCTTAAACGAGTATGTGAGCATCTCAATCTTTTGGAAAGA 917

135 CTATTTGGCCTCGAGTTTCTGTGATCACAAAAGATCACGGTGTGGTCTCCTAA 194

Db 918 CTATTTGGTCTAGCCATTGGGATAACGCAACCTCTAAGACATGGCTGGATTCGCGCAA 977

OV 195 ACCCATTTGTGAAACAGATTAGAAGGCCAAAGCACGTTGTTGTTAAGTTTGTGGTGAAATT 254

Db 978 AGAAATAAAAAAGCAGGTTCTGTGG--TGTCCTTGGAAATTTACATTTAAAGTAAAGTT 1034

QY 255 CTTTCGGCCTGACCCACACACAACCTCCAAGAAAGAACTCACAAAGGTACCTGTTCGGCGCTGCA 314

Db 1035 TTATCCACCTGACCCAGCACAGTTAACAGAAGACATAACAAGATATTTATGTCTTCA 1094

QY 315 GGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACCAGCCAGCTCTCTT 374

Db 1095 GCTTCGGCAGGACATAGTTCAGGACGCTGCCCTGTTCCTTGC AACCTTAGCATTT 115

QV 375 GATTTACACACATTGTGCAATCTGAGATTGGGGATTTTGATGAAGCCTTGGACAGAGAGCA 434

Db 1155 AGGTTCTTACACCATCCAGTCTGAACTGGGAGACTACGACCCAGAACTCCATGGCGTGGA 1214

QY 435 CTTAGCAAAAATAAATACATACC-----TCAGCAAGACGCACCTAGAGGACAAAAATCGT 488

Db 1215 TTATGTTAGTGATTTTAAACTGGCCCCGAATCAGACCCAAGGAAC TTGAAGAGAAGGTCAT 1274

QY 489 GGAATTTACCATTAACCAACATTTGGACAAACACCAAGCAATCAGATTTCCAGCTCCTAGA 548

Db 1275 GGAAC TGCATAAGTCATACAGGTCCAGCTCCAGCTCAGGCTGACTGGAGCTTCTTGA 1334

QY 549 GATTGCCCGTCGGCTAGAGATGTATGGAATCCGGTTGCACCCGGCCAAAGGACAGGGAAGG 608

Db 1335 GAATGCCAAAAGTGTCTATGTATGGAGTTGATCTTCATAAAGCAAAGGACTTGAAGG 1394

QY 609 CACGAAGATCAATCTGGCCGTTGCCAACACGGGAATTCTAGTGTTCAGGGTTTCACTAA 668

Db 1395 AGTAGATATCATCCTAGGTGTCTGCTCTAGTGGCCCTTCTGGTTTACAAAGATAAGCTGAG 1454

QY 669 GATCAATGCCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTTCT 728

D	b	1455	AATTACCGCTCCCTTGGCCCCAAAGTGCTGAAGAATTTCTTTATTAACAAGTAGTACTGTTTT	1514
Q	y	729	CATCAAGCTCCGCCAGATGCCAATAGTCGTACCAGGATAACCTTGGAAATTCCTGATGGC	788
D	b	1515		1574
Q	y	789	CATCGGGGATTTCTGCAAGTCTCTTCGGAAATCTGTGTGAACATCATGCTTCTTTAG	848
D	b	1575		1634
Q	y	849	ACITTTTGAAGAGCCCCAACCAAGCCCAGCCCTCTCTTTAGCCGGGGTGCATCAT	908
D	b	1635	A-----TTGACATCTACACAGACCATTTCCCAAAGCAAAATTTCTTGGCTPAGGATCCAAT	1691
Q	y	909	TGCGTTCAGTGGTCGCACTCAGAAGCAG	936
D	b	1692		1719
RESULT 9				
I	D	ABK84074		
X	X	ID ABK84074 standard; cDNA; 3643 BP.		
X	X	ABK84074;		
X	X	XX		
X	X	14-AUG-2002 (first entry)		
X	X	Human cDNA differentially expressed in granulocytic cells #645.		
K	W	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
K	W	viral infection; parasitic infection; protozoal infection;		
K	W	fungal infection; sterile inflammatory disease; psoriasis;		
K	W	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
K	W	cardiac reperfusion injury; renal reperfusion injury; ARDS;		
K	W	adult respiratory distress syndrome; inflammatory bowel disease;		
K	W	Crohn's disease; ulcerative colitis; periodontal disease;		
K	W	granulocyte activation; chronic inflammation; allergy.		
X	X	XX		
O	S	Homo sapiens.		
X	X	XX		
P	N	WO200228999-A2.		
X	X	XX		
P	D	11-APR-2002.		
X	X	XX		
P	F	03-OCT-2001; 2001WO-US30821.		
X	X	XX		
P	R	03-OCT-2000; 2000US-237189P.		
X	X	XX		
P	A	(GENE-) GENE LOGIC INC.		
X	X	XX		
P	I	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
X	X	XX		
D	R	WPI; 2002-435328/46.		
X	X	XX		
P	T	Detecting granulocyte activation by detecting differential expression		
P	T	of genes associated with granulocyte activation, which serves as		
P	T	diagnostic markers that is useful for monitoring disease states and		
P	T	drug toxicity -		
X	X	XX		
P	S	Claim 1; SEQ ID NO 645; 114pp; English.		
X	X	XX		
C	C	The invention relates to detecting (M1) granulocyte (GC) activation		
C	C	(GCA), by detecting the level of expression of gene(s) (Gs) identified		
C	C	DNA chip analysis as given in the specification, and comparing		
C	C	the expression level to an expression level in an unactivated		
C	C	GC, where differential expression of Gs is indicative of GCA.		
C	C	Also included are modulating (M2) GA by contacting GC with an agent		
C	C	that alters the expression of at least one gene in Gs; (2) screening		
C	C	for an agent capable of modulating GCA or an inflammation (especially		
C	C	chronic) in a tissue, an allergic response in a subject, exposure of		
C	C	a subject to a pathogen or sterile inflammatory disease using the		
C	C	gene expression profile; (3) detecting (M4) an inflammation (especiall		
C	C	chronic) in a tissue, an allergic response in a subject exposed to a		

subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating CA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

```

xx
SQ      Sequence 3643 BP; 1038 A; 815 C; 863 G; 927 T; 0 other;

Query Match          7.7%;   Score 181.8;   DB 24;   Length 3643;
Best Local Similarity 51.9%;   Pred. No. 1.6e-40;
Matches 514; Conservative 0; Mismatches 462; Indels 15; Gaps 4;

QY      24  CAAATCCAGATGCTGTACACCCAGGAGCATTTGAAGTTCACAAAAGAGCTCCTGG 83
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DY      864  CAACATCCTTCTTGATTAACACTGTACAAGCTTTCAAAGTCAATAACATGATCAGG 923
QY      84  GAAGGTGCTGTGCATGCAGTTTGCACCACCTCAACCTCGTGGAGGTGACTATTFTGG 143
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      924  GCAAGTCTGTGGATGTCGTCTTCAAGCATCTAGATTTGACTGAGCAGGACTATTFTGG 983
QY      144  CCTCGAGTTTCCCTGATCA--CAAAGAATCACGGTGTGGCTGCATCTCTA AACCCAT 200
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      984  TTTCAGCTTGCTCGTATGCCACAGATAACCCAAAGTGGCTGGATCCAACAACCAAT 1043
QY      201  TGTCAAACAGATTAGAAGGCCAAAGCACGTTGTGTAAAGTTTGTGTGAAATCTTTTCC 260
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1044  AAGGAGCAGCTAAGAGAGGATCTCCTTACAGTTTGAAGTTAGATCAAAATTTTGT 1103
QY      261  GCCTGACCACACAACTCCCAAGAAGAACTCAAGGTACCTGTTCCGCTGCAGGTGAA 320
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1104  AAGTGACCCCCAACAGTTACAAGAAGAATATACAAGGTACCAGTATTTTTTGCAAAATA 1163
QY      321  CGAGACTTGCTCAAGGCAGTTTGACGTGTAATGACACAGCCAGCTCTCTTGATTC 380
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1164  ACAAGACATCTTACTCGAAGATTACCCCTGCTCTTAATACTGCTGCCCTTTTAGCTTC 1223
QY      381  ACACATTGTGCAATCTGAGATTGGGGATTTTGTAGTGAAGCCTTGGAC--AGAGAGCACTT 437
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1224  ATTTGTGTTTCAGTCTGAACCTTGGAGACTAGCATCAGTCAGAGAAGCTTGTACAGCTACCT 1283
QY      438  AGCAAAAAATAATACATACCTCAGCAA--GAGCGCACTAGAGACAAATACTGCGGAATT 494
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1284  CTCAGATTATCTTTTCATCTCCTAATCAACCTCAAGATTTTGAAAAGAAATTCGCAAAAT 1343
QY      495  TCACATAACCAATATGGACAACACAGCAGAGATCAGATTTCCAGCTCTCTAGAGATTGC 554
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
DB      1344  ACATCAGCAACACATAGGCTTATCTCCTGCAAGACAGAAATTTAATACCTAAACACAG 1403
QY      555  CCSGTGGCTAGAGATGTATGGAATCCGGTTCACCCCGCCAGGACAGGAGGACGCA 614
DB      ||| |||| | |||| | |||| | ||| |||| | ||| |||| | ||| |||
DB      1404  ACGTACCTTAGAATCTATGGAGTTGAATTCACCTATGCAAGGGATCAGAGTAACAATGA 1463
QY      615  GATCAATCTGCGCCCTTGCCAAACAGCGGAATTTCTAGTGTTCAGGCTTTTCACTAAGATCAA 674
DB      ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
```


QY 629 TTGCCAACACGGGAATCTAGTGTTCAGGGTTTCACCTAAGATCAATGCCTTCAACTGGG 688
D 1066 TTTGTGCAATGGCTGCTCATCTACCGGACCGGCTGAGATCAACCGCTTGGCTGGC 1125
QY 689 CCAAGGTGGGAAGCTGAGCTTCAAGAGAGAGCGCTTTCTCATCAAGCTCCGGCGAGATG 748
D 1126 CCAAGATCCTCAAGATCTCCTACAAAGAGAGTAACCTTCTATATCAAGATCGGGCTGGG 1185
QY 749 CCAATAGTGGCTACAGGATACCTTTGGAATTCCTGATGGCCAGTCGGGATTTCTGCAAGT 808
D 1186 AGTATGAGCAATTTGAGAGCAATTTGGCTTTAAGCTCCCAACACCGGTCAGCCAAAGA 1245
QY 809 CTTCTGGAAATCTGTGTTGAACATCATGCTCTTTTAGACTTTTGAAGAGCCCAAC 868
D 1246 GACTGTGGAAGTCTGATCGAGCAATACATTTCTCCGGC---TGTGTCCCTGAGC 1302
QY 869 CAAAGCCCAAGCCGCTCTCTTTAGCCGGGGGTCAATATTTCCGGTTCAAGTGGGACTC 928
D 1303 CCCCACCCCAAGGGCTTCTGGTGA---TGGGCTCCAAGTTCCGGTACAGTGGGAGGACCC 1359
QY 929 AGAAGCAGGTTCTCGA 944
D 1360 AGGCACAGACTCGCCA 1375

RESULT 11

ABL12469
ID ABL12469 standard; cDNA; 2916 BP.

XX AC ABL12469;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31889.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB68366.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX PS Claim 1; SEQ ID NO 31889; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_sequences.

XX SQ Sequence 2916 BP; 833 A; 789 C; 718 G; 576 T; 0 other;
Query Match 7.38; Score 171.4; DB 23; Length 2916;
Best Local Similarity 51.0%; Pred. No. 1.3e-37;
Matches 462; Conservative 0; Mismatches 31; Indels 12; Gaps 2;

QY 35 TGCCTGGATGACACCCAGGAGGCATTTGAAGTTCCACAAGAGCTCCTGGGAAGGTGCTGC 94
D 486 TGCCTGGATGGCTCCCTCTGACGCTGTCATGATCGCAAGCCATTTGGCCGTGATGGA 545
QY 95 TGGATGCTAGTTTGAACAACACCTCAACCTCGTGAAGTGAATTTTGGCCCTCGAGTTTC 154
D 546 TCAACTCAATTCGCGCGTCTGAACCTCATCGAAAAGGACTACTTTTGGTGTGACCTATG 605
QY 155 CTGATCACAAAAGATCACGGTGTGGCTGGATCTCTTAAACCCCATTTGTGAACAGATT 214
D 606 AGAGCCCAACAGATCCCGGCACCTTGGCTGGATCTGGGAAGCCGCTATCCAAATCTTTC 665
QY 215 GAAGCCCAAGACAGCTTGTGTTAAGTTTGTGTTGAAATTTCTTCCGCTGACCAACAC 274
D 666 GCACGGACACTT---GGCCTCTACCTTCGCGCTCAAGTTCTATCCCGCGGAGCCATCGC 722
QY 275 AACTCCAAAGAAAGAACTCACAAAGTACCTTTCGCGCTGCAAGTGAAGCAGGACTTGG 334
D 723 AGCTGAAGGAGGACATCAGCGCTACCATTTGTGCTGCAGGTGCGCAATCAGCATCTCTG 782
QY 335 AAGCAGGTTGACGTGTAATGACACAGCGAGCTCTCTTGTGATTTACACATTTGTGCA 394
D 783 AGGCTGACTGCCCTGCACATTCCTCACCCACGCTCTGCTCGGATCTTACTTGTGTGAG 842
QY 395 CTGAGATTGGGGATTTTGATGAAG-----CCTTGGACAGAGAGCACATTAGCAAAA 445
D 843 CGGAGATGGGCGACTACGATGCAGAGAAATGCCACCAGGCGCTACCTGAAGGACTTCA 902
QY 446 ATAATATACATACCTCAGCAAGACGCACCTAGAGGACAAATCGTGAATTCACCATAA 505
D 903 AGATCGCTCCAAATCAGACGGCTGAGTTGGAGGATAAGGTCATGATCTTCAACAAGACC 962
QY 506 ACATTGCAACAAACAGCAGCAATCAGATTTCCAGCTCTCAGATATGCCCGTGGCTAG 565
D 963 ACAAGGGACAAATCGCCCGCGAGGCTGAGCTACACTACCTGGAGAAATGCCAAGAGCTG 1022
QY 566 AGATGATGAATCCGGTTGCACCCGCGCAAGGACAGGAGGACGAGATCAATCTGG 625
D 1023 CCATGTACGGCGTGACTTGCTATCCCGCTAAGGATTTGAGGGCGTGACATCATGCTGG 1082
QY 626 CCGTTGCCAACACAGGGAATTTCTAGTGTTCAGGGTTTCTACCTAAGATCAATGCCTTCA 685
D 1083 CGTTTGTGCTCCGGTTTGTCTCTACCGCGATAAATTCGGCATCAACCGCTTGGCT 1142
QY 686 GGGCCAGGTGCGGAAGCTGAGCTTCAAGAGAGAGCGGTTTCTCATCAAGTCCGGCCAG 745
D 1143 GGGCCAGATTTCTGAAGATCTCTTACAGCGCCACCATTTCTATACATCAAGATCCGAC 1202
QY 746 ATGCCAATAGTGGCTTACAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCT 805
D 1203 GTGAATTCGAGCAGTACGAATFCCACCATTTGGCTTTAAGCTGGCCCAACCATCGTGG 1262
QY 806 AGTCCCTTCTGGAAAATCTGTGTTGAACATCATGCTCTTTTAGACTTTTGAAGAGCCCA 865
D 1263 AGAAATGTGGAATCTCTGCTGGAGCACCACACTTCTTCCGCTGATGATCCCCCGAAC 1322
QY 866 AACCAAGCCCAAGCCGCTCTCTTTAGCCGGGGGTATCATATTTGGGTTAGTGTGCGGA 925
D 1323 CGGTCAAGCAAGTCCAAGATGTTCCAGTCTTTCGGGTCCACCTATCGCTACAAGGGT 1382
QY 926 CTCAG 930
D 1383 CTCAG 1387

RESULT 12

ABL18111
ID ABL18111 standard; DNA; 3470 BP.

AC ABL18111;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5806.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Claim 1; SEQ ID NO 5806; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (AB01840-ABL16175) and the encoded proteins (AB857737-AB872072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3470 BP; 974 A; 958 C; 883 G; 555 T; 0 other;

XX Query Match 7.3%; Score 171.4; DB 23; Length 3470;

XX Best Local Similarity 51.0%; Pred. No. 1.4e-37;

XX Matches 462; Conservative 0; Mismatches 431; Indels 12; Gaps 2;

QY 35 TCCTGTGATGACACCCAGGAGGCAATTGAAGTCCACAAAGAGCTCTGGAGGTCCTGC 94

DB 482 TGTGTGATGCTCCCTTCTGGAGCTCTCCATTCATCGCAAGGCAATGGCGGTGATGTA 541

QY 95 TGGATCAGTTTCAACACCACTCAACTCGTGAAGGTGACTATTTTGGCTCGAGTTTC 154

DB 542 TCNACTCAATCTGCGCGGTCTGCACTTCATCGAAAGGACTACTTTGGTCTGACCTATG 601

QY 155 CTGATCACAAAAGATACACGGTGTGGCTGATCTCTTAAACCCATTGTGAACAGATTA 214

DB 602 AGACGCCACAGATCCGCGCACTTGGCTGGATCTGGAGAAGCGGTATCCAAGTTCTTC 661

QY 215 GAGGCCAAGACAGCTTTGTTAACTTTGTGGTGAATTTCTTCGCTGACCAACAC 274

DB 662 GCACGACACTT---GSCCTCTACCTTCGCGTCAAGTTCTATCGCGGAGCCATCGC 718

QY 275 AACTCCAAGAAGAACTCAAAAGTACTCTGTCGCGTGCAGGTGAAGAGGACTTGGCTC 334

DB 719 AGCTGAAGGAGGACATACGCGCTACCATTTGTGCTGCGAGGTGCGCAATGACATCCTGG 778

XX Venter JC, Adams M, Li PWD, Myers EW;

QY 335 AAGCAGGTTGACGTGTAATGACACACGCGCAGCTCTCTTGTATTTTACACATTTGTCAAT 394
DB 779 AGGTGCGACTGCGCTGCATTCGTCAACCCAGCTCTGCTCGGATCTTACTTGGTCAGT 838
QY 395 CTGAGATTGGGGATTTTGTATGAAG-----CCTTGGACAGAGCAGCTTAGCAAAA 445
DB 839 CGGAGATGGCGGACTAGCATGCGAGAGAAATGCCACCAGGGCTTACCTGAAGGACTTCA 898
QY 446 ATAAATACATACCTCAGCAAGACGCACTAGAGGACAAAATCGTGGAAATTTCCACATAAC 505
DB 899 AGATCCCTCCAAATCAGACGGCTGAGTTGGAGGATAAAGTCATGGATCTTCAACAACCC 958
QY 506 ACATTGGCAAAACACACAGCAGAAATCAGATTTCCAGCTCCTAGAGATTGCCCGTGGCTAG 565
DB 959 ACAAGGCAAAATCGCCCGCGAGGCTGAGCTACACTACCTGGAGAAATGCCAAGAAGCTGG 1018
QY 566 AGATGATGAATCCGGTTGCACCCGCGCAAGGACAGGGAAGGACGAGATCAATCTGG 625
DB 1019 CCATGTACGGCGTGGACTTGCATCCCGCTAAGGATTTCTGAGGCGGTGGACATCATCTGG 1078
QY 626 CCGTTGCCAACACGCGGAATTTCTAGTGTTCAGGGTTTCACTAAAGATCAATGCCCTTCAACT 685
DB 1079 GCGTTTGTGCTCCGGTTTGTCTGTCTACCGCGATAAATTCGCGATCAACCCGTTTGCCT 1138
QY 686 GGCCAAAGGTGGGAAAGCTGAGCTTCAAGAGGAAGCGCTTTCTCATCAAGCTCCGGCCAG 745
DB 1139 GGCCAAAGATTTCTGAAGATCTCCTCAAGCGCCACCATTTCTTACATCAAGATCCGACCG 1198
QY 746 ATGCCAATAGTGGTACCAGGATACCTTGGAAATTTCTGATGGCCAGTCGGGATTTCTGCA 805
DB 1199 GTGAATTCGAGCAGTAGCAATCCACATTTGGCTTTAAGTGGCCAAACCATCTCGTGGCCCA 1258
QY 806 AGTCCTTCTGGAAAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCA 865
DB 1259 AGAACTGTGGAATCTCTGCTGGAGCACCACACCTTCTTCCGCTGATGACCCCGCAAC 1318
QY 866 AACCAAGCCCAAGCCGCTCTTTAGCGGGGGTGCATCATTTTCGGTTCAGTGGTCGGA 925
DB 1319 CGGTACGCAAGTCCAAGATGTTCCAGTCTTCCAGTCTTCCGGTCCACCTATCGTACAAGGTCGCA 1378
QY 926 CTCAG 930
DB 1379 CTCAG 1383

RESULT 13

ABL18211

ID ABL18211 standard; DNA; 5897 BP.

XX AC ABL18211;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6106.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX


```
Db 9156 GGAGTGAGTGGCTGTGTCTTTGGGAGGTGGGTAGCCCTGGAATTTTGCTTCACAG 9215
QY 248 TGAATTCCTTCGGCTGRACCACACACACTCCAAGAAGACTCACAAGGTACTGTTGG 307
Db 9216 TCAAGTTCCTACCGCGCTGTATCCCTGCCAGCTGACAGAAGACATCACAAGATACTACTGT 9275
QY 308 CGCTGCAGGTGAAGCAGGACTTGGCTCAAGGCAGGTTGACGTGTAATGACACCAGCGCAG 367
Db 9276 GCTGCAGCTGGGGCAGACATCATCAGGGCGGCTGCCATGCTCTTTGTCACGCATG 9335
QY 368 CTCTCTGTATTTACACATTTGCAATCTGAGATTTGGGATTTTGATGAAGCCTTGGACA 427
Db 9336 CCCTACTGGGCTCTCTACGCTGTGCAGGCTGAGCTGGGTGACTATGATGCTGAGGAGCATG 9395
QY 428 GAGAGACCTTAGCAAAAATAAATACATACC-----TCAGCAAGAGCAGCTAGAGGACA 481
Db 9396 TGGGCAACTATGTACAGCGAGCTCCGCTTCGCCCTTAACCCAGACCCGGGAGCTGGAGAGA 9455
QY 482 AAATCGTGAATTTACCATTAACCATGGACAAACACCCAGCAGAAATCAGATTTCCAGC 541
Db 9456 GGATCATGGAGCTGCATAGACATATAGGGGGATGACCCCGGAGAGCAGAAATCCACT 9515
QY 542 TCCTAGAGATTCGCCCTCGGCTAGAGATGTATGGAATCCGGTTGCACCCGGCCOAGGACA 601
Db 9516 TCTTAGAATGCCAAGAAGCTTTCCATGTACGGAGTAGACCTGCACCATGCCAAGGACT 9575
QY 602 GGAAGGCAGAGATCAATCTGGCCGTTGCCAACACAGGGGAATTCAGTGTTCAGGGTT 661
Db 9576 CTGAGGCGATCGACATCATGTTAGGCGTTTGTGCCAATGGCCCTGCTCATCTACCCGGACC 9635
QY 662 TCACTAAGATCAATGCTTCAACTGGGCCAAGGTGCGGAAGCTGAGCTCAAGAGGAAGC 721
Db 9636 GGCTGAGAATCAACCGTTTGCCCTGSCCCNAGATCCTCAGATCTCTACAGAGAGTA 9695
QY 722 GCTTTCATCAAGCTCCGGCCAGATGCCAATAGTCGCTACCAAGGATACCTTGAATTC 781
Db 9696 ACTTCTATATCAAGATCCGGCCTGGGGAGTATGAGCAATTTGAGAGCACAAATGGCTTTA 9755
QY 782 TGATGCCAGTCGGGATTTCTGCAAGTCTCTTGAAATCTGTGTGAACATCATGCGCT 841
Db 9756 AGCTCCCAAAACCCAGCTAGCCCAAGAGACTGTGGAAGTCTGCATCGAGCATACAT 9815
QY 842 TCTTTAGACTTTTGAAGAGCCCAAAACCAAGCCCAAGCCGTCCTCTTTAGCCGGGGT 901
Db 9816 TCTTCGGC---TGGTGTCCCTTGAGCCCCACCCAGGGCTTCCCTGGTGA---TGGGCT 9869
QY 902 CATCATTCGGTTTCAGTGTGCGACTCAGAAGCAGGTTCTCGA 944
Db 9870 CCAAGTTCGGGTACAGTGGGAGGCCCCAGGCACAGACTCGCCA 9912
```

Search completed: December 3, 2002, 18:24:07

Job time : 476.636 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 71.4684 Seconds
(without alignments)
10088.332 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_2501
Perfect score: 2351
Sequence: 1 ccttcaggaaactcgtgc.....attgaggagcggaagacga 2351

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.8	2.4	4080	US-08-446-345-35	Sequence 35, Appl
2	53.8	2.3	7218	US-08-232-463-14	Sequence 14, Appl
3	42.2	1.8	4411529	US-09-103-840A-1	Sequence 1, Appl
4	40.8	1.7	4776	US-08-852-401-1	Sequence 1, Appl
5	39.6	1.7	4722	US-08-979-608A-14	Sequence 14, Appl
6	38.4	1.6	4267	US-08-949-155-51	Sequence 51, Appl
7	38.4	1.6	4267	US-09-819-964-51	Sequence 51, Appl
8	38.2	1.6	1053	5352575-6	Patent No. 5352575
9	37.6	1.6	2619	US-08-467-822-19	Sequence 19, Appl
10	37.6	1.6	2619	US-08-432-697-19	Sequence 19, Appl
11	37.6	1.6	2619	US-08-466-248-19	Sequence 19, Appl
12	37.2	1.6	562	US-09-404-879A-147	Sequence 147, App
13	37.2	1.6	44377	US-08-804-227C-7	Sequence 7, Appl
14	37.2	1.6	44377	US-08-804-198-1	Sequence 1, Appl
15	36.4	1.5	2824	US-09-010-928B-3	Sequence 3, Appl
16	36.2	1.5	822	US-09-228-986-38	Sequence 38, Appl
17	36.2	1.5	4403765	US-09-103-840A-2	Sequence 2, Appl
18	36	1.5	2214	US-08-864-038A-1	Sequence 1, Appl
19	36	1.5	3331	US-08-864-038A-2	Sequence 2, Appl
20	36	1.5	3331	US-08-864-038A-4	Sequence 4, Appl
21	36	1.5	56516	US-08-996-306-1	Sequence 1, Appl
22	36	1.5	56516	US-09-338-907-1	Sequence 1, Appl
23	36	1.5	56516	US-09-218-207-1	Sequence 1, Appl
24	36	1.5	56520	US-09-338-907-179	Sequence 179, App
25	36	1.5	56520	US-09-218-207-179	Sequence 179, App
26	35.4	1.5	2132	US-09-552-322-1	Sequence 1, Appl
27	35.4	1.5	4137	US-09-499-964-2	Sequence 2, Appl

28	35.4	1.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
29	35.4	1.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl
30	35	1.5	289	4	US-09-007-005-17	Sequence 17, Appl
31	35	1.5	289	4	US-09-244-796-17	Sequence 17, Appl
32	34.8	1.5	4822	4	US-09-484-970B-106	Sequence 106, App
33	34.6	1.5	657	4	US-09-527-345-3	Sequence 3, Appl
34	34.6	1.5	1029	3	US-08-911-853-1	Sequence 1, Appl
35	34.6	1.5	1029	4	US-09-479-409-1	Sequence 1, Appl
36	34.6	1.5	1029	4	US-09-479-453-1	Sequence 1, Appl
37	34.6	1.5	2040	3	US-08-604-789B-1	Sequence 1, Appl
38	34.6	1.5	2040	3	US-08-604-789B-12	Sequence 12, Appl
39	34.6	1.5	2040	4	US-09-312-721A-1	Sequence 1, Appl
40	34.6	1.5	2040	4	US-09-312-721A-12	Sequence 12, Appl
41	34.6	1.5	4377	3	US-08-911-853-28	Sequence 28, Appl
42	34.6	1.5	4377	4	US-09-479-409-28	Sequence 28, Appl
43	34.6	1.5	4377	4	US-09-479-453-28	Sequence 28, Appl
44	34.6	1.5	10348	2	US-08-457-273B-41	Sequence 41, Appl
45	34.6	1.5	10348	3	US-08-556-419-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-446-345-35
; Sequence 35, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASES PTP-D1
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; Zip: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-446-345-35

Query Match 2.4%; Score 56.8; DB 2; Length 4080;
Best Local Similarity 46.6%; Pred. No. 3.7e-06;
Matches 218; Conservative 0; Mismatches 247; Indels 3; Gaps 1;

Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;
QY 1800 GAAGCACCTGGCGGCTACCTGTGGAAGACACAGCGAGCCCTGGAGATGG 1859
DB 618 GGAGAAGCTGGCGGCTCTGTGCAAGAAGTATGCGGAATCTGTCAGGAGCACCGGAATC 677
QY 1860 AATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAAGAGTGTG 1919
DB 678 GCAGAAGCAGATGAAGCTGCTGCAGAAGAAGCAGAGCCAGCTGGTGCAGAGGAAG-- 733
QY 1920 TTACCTACCGCTCAACACCTTCTCTCTGGGCACTGACCGGCTCATGCACTACAAGCA 1979
DB 734 --ACCACCTGGCTGGCGAGCAGCAGCAAGGCCATCTCTGGCCGCGCAGCAAGCTCGAGAGCT 791
QY 1980 GTCTCTGGAGCGGCTGTGCAAAACACCCCGCGAGCCAGCCGACTTCAGGAGTGGCG 2039
DB 792 GTGCCGAGGAGTGCAGCGGCACACCCGCTCGCTCAGGAAGAAGTGTGCGAGCGAGCCG 851
QY 2040 AGCCGCTTTGGCAGAGATCAGGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGAT 2099
DB 852 AGAGGAGGAGGAAGCAAGGAGGTGACGTACACACTTCCAGATGACGCTCAAGCAT 911
QY 2100 GGAGAATTTCCAGAGCTGCAGGA 2123
DB 912 TCAGCTGCAGATGGAGCAGCAAA 935

RESULT 6

US-08-949-155-51
; Sequence 51, Application US/08949155
; Patent No. 6271436
; GENERAL INFORMATION:
; APPLICANT: Piedrahita, Jorge A
; APPLICANT: Bazer, Fuller W
; TITLE OF INVENTION: Compositions and Methods for the
; Generation of Transgenic Animal Species
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/949,155
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/027,338
; FILING DATE: 11-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,094
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: TAMK.177
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-949-155-51

Query Match

Best Local Similarity 47.2%; Pred. No. 0.82;
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1225 GCCGCCTGGCGCCACCGAGGAGAGAGGAGGTGCTTAAGATAGACCCAGCAGAGT 1284
DB 3020 GCCCTCTCTGGCGCCAGGAGCTGATAGAGGAGAGCATGAAGAGGTGAAGCCCTACCGC 3079
QY 1285 AAACCTCAGCCCCCGCAGCCCAAGCAGGCTCCCTGACTGGCAGTCCCTCAGCTTTCCGAG 1344
DB 3080 GAGGAGCTGGAGGCGCAGCTGGGCCCGGTGACCCAGGAGAGCGAGCGCGCTGTCCAAG 3139
QY 1345 CTGTCTGTGAACCTGCGAGGGGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCCAACCTG 1404
DB 3140 GAGCTGAGGCGCGCAGGCCCGCGTGGCGCCGCACATGGAGGAGCGTGGCAACCGCTTG 3199
QY 1405 AGCCCCGACACAAAGCAGGCGCTCTCCCTTATGATCAGCCGCGCTGCTGAATGACAGGCGTCG 1464
DB 3200 GTGCTTACCGCAGCGAGGTGCAACAACATGTTGGGCCAGACCCAGGAGGAGCTGCGGAGC 3259
QY 1465 CCCGGAC 1472
DB 3260 GCCTGGC 3267

RESULT 7

US-09-819-964-51
; Sequence 51, Application US/09819964
; Patent No. 6369294
; GENERAL INFORMATION:
; APPLICANT: Piedrahita, Jorge A
; APPLICANT: Bazer, Fuller W
; TITLE OF INVENTION: Compositions and Methods for the
; Generation of Transgenic Animal Species
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE AND DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/819,964
; FILING DATE: 28-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,155
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 60/046,094
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: TAMK.177
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-819-964-51

QY 505 CACATTGGACAACACAGCAATCAGATTTCCAGCTCCTAGA 548
 | | | | | | | | | | | | | | | | | | | | | | | |
 Db 445 GCGATCGTCTGTCAGGTGGGATCACATTTCCACTTCTCGA 488

RESULT 10

```

US-08-432-697-19
; Sequence 19, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence."
US-08-432-697-19

```

Query Match 1.6%; Score 37.6; DB 4; Length 2619;
 Best Local Similarity 51.8%; Pred. No. I.I.;
 Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 385 ATTGTGAATCTGAGATTGGGATTTGATGAAGCCTTGGACAGAGACACTTAGCAAAA 444
 | | | | | | | | | | | | | | | | | | | | | | | |
 Db 325 ACTATCCACACTCCGCTAGAGATATGCAAAATAGCCCCCGGCGAGGTCTTCTTAAA 384
 | | | | | | | | | | | | | | | | | | | | | | | |
 QY 445 AATAATACATACCTAGCAAGACGCACTTAGAGGACAAAATCGTGGAAATTTCCACCATAC 504
 | | | | | | | | | | | | | | | | | | | | | | | |

Db 385 AATGAGACATTACTTAACCGCGCAAGAGCCATTAGCTTGAAGTGAATAATAA 444
 QY 505 CACATTGGACAACACAGCAATCAGATTTCCAGCTCCTAGA 548
 | | | | | | | | | | | | | | | | | | | | | | | |
 Db 445 GCGATCGTCTGTCAGGTGGGATCACATTTCCACTTCTCGA 488

RESULT 11

```

US-08-466-248-19
; Sequence 19, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
; OTHER INFORMATION: sequence."
US-08-466-248-19

```

Query Match 1.6%; Score 37.6; DB 4; Length 2619;

Best Local Similarity 51.8%; Pred. No. 1.1;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 385 ATTGTGCAATCTCAGATTGGGGATTGATGAAGCCTTGACAGAGACGACCTTAGCAAAA 444
Db 325 ACTATCCACACTCCGGTAGAGGATAATGGCAATTAGCCCCCGCGAGGTCTCTTTAAAA 384

QY 445 AATAAATACATACCTCAGCAAGACGACAGAGACAAAATCGTGAATTTCCACCATAA 504
Db 385 AATGAGACATTACTATTACGCGCGCAAGAGCCATTAGCTTGAAGTGAAATATAA 444

QY 505 CACATTGGACAACACGACAGATCAGATTTCCAGCTCTAGA 548
Db 445 GCGATCGCTCTGTCAGTGGGATCACATTTCCACTTCTCGA 488

RESULT 12
US-09-404-879A-147
; Sequence 147, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-147

Query Match 1.6%; Score 37.2; DB 4; Length 562;
Best Local Similarity 53.4%; Pred. No. 0.58;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 43 GACACCCAGAGGCATTGAAGTTCACAAAGAGCTCTCGGGAAGTGCTGTGGATGCA 102
Db 175 GATGCAGAGCTGGAGTTTGCAATCCAGCCAAATACAACTGGAAACAGCTTTTGTATCAG 234

QY 103 GTTGTCAACCACTCAACCTCGTGGAGGTGACTATTTTGGCTCGAGTTCTCTGATCAC 162
Db 235 GTGGTAAGACTATCGGCTCCGGGAGTGTGTTACTTTGGCCCTCCACTATGTGGATAAT 294

QY 163 AAAAAGATCACGGTGTGGTGATCT 188
Db 295 AAAGGATTTCCTACCTGGCTGAAGCT 320

RESULT 13
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 1.6%; Score 37.2; DB 2; Length 44377;
Best Local Similarity 43.8%; Pred. No. 6.9; Mismatches 208; Indels 0; Gaps 0;
Matches 162; Conservative 0;

QY 1729 GGCCGCTCAATGCCCCAAATCAGAGATTACCAAGAATCGGGATGTCATGCTGAAGAAC 1788
Db 29515 GACTTCGCGACCCGCTTCACCGGAGCGGACGACGTCGCCGCGGACGTCGCTGAAC 29574

QY 1789 ATTCAGGATGAAGCACCTGGGGCTCACCTGTGGAAGCACAGCAGGCGCTTTGGAGGCC 1848
Db 29575 TCGCTCACCAAGAGTTCGTGGACGCTCCCTCGGGCTGCTCGGCGCGCGCTTC 29634

QY 1849 CTGGGAATGGAATCAAGACTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 1908
Db 29635 CTGGAGCTGGCAAGACCGACGTCGCGGACCCGCGGAGCGGATCGCCCGCAACACCCCGG 29694

QY 1909 CAGAAGTGTGTACCTACCGCTCAACACACTTCCTCTCGGGCCACTGCACCGGCTCATG 1968
Db 29695 GTGCGTACCGGGCGTTCACCTCAACAGAGCGGCGGACGACACTCGCGCGCTGCTG 29754

QY 1969 CACTACAAGCAGTCTCTGGAGCGGCTGTGCAACACACCCCGCGGAGCCACCGGCTTC 2028
Db 29755 CGGGAATGATGACCTGTTTCGCGCGGCGGTGCTCACCGGCTGCCGCTCGTCAACCCAC 29814

QY 2029 AGGAGTCCGAGCGGCTTTGGCAGAGATCAGGAGATGTTGGCAGAGCTCCACCGTACG 2088
Db 29815 GACGTGCGCGCGCGGCGGACGCTTCGCGCACCACATCAGCAGGCGCGGACACCCGGAAG 29874

QY 2089 ATGATCAAGA 2098
Db 29875 CTCGTCCTGA 29884

RESULT 14
US-08-804-198-1

Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rosteck, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match 1.6%; Score 37.2; DB 2; Length 44377;
Best Local Similarity 43.8%; Pred. No. 6.9; Mismatches 0; Indels 0; Gaps 0;
Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 1729 GCGCGCTAAATGCCCAATACAGATTACCAAGAATCGCGGATGTCATGCTGAAGAAC 1788
DB 29515 GACTTCGCGACCGCTTACCGGAGCGGACGCGCTCCCGGCGGACGTCGCTCTGAAC 29574
QY 1789 ATTGAGGCGATGAAGCACCTCGGGGCTACCTGTGAAGCAGCAGGAGCCCTTGAGGGCC 1848
DB 29575 TCGCTCACCAGAGTTCGTGGACGCTCCCTCGGGCTGCTCCGTCGCGGCGCGGTTTC 29634
QY 1849 CTGGAAGTGAATCAAGAGCTCCCGGCGCTGGAGACTTCTGAGAGACTTTGAGCTG 1908
DB 29635 CTGGAGCTGGGCAAGCAGCAGCTCCGGGACCCGAGCGGATCGCCGCGGCAACACCCCGG 29694
QY 1909 CAGAAAGGTGTGTACCTACCGCTCAACACCTTCCCTCCTCGGGCCACTGCACCGGCTCATG 1968

DB 29695 GTGCGCTACCGGGCGTTCGACCTCAACGAGGCGCGGACGACTCGGCGGCTGCTG 29754
QY 1969 CACTACAAGCAGGCTCTGGAGCGGCTGTGCAAAACACACCCCGGAGCAGCCGACATTC 2028
DB 29755 CGGGAAGTGAAGGACCTGTTCGCCGCGCGTGTGTCACCCGCTGCCGCTGCTCACCCAC 29814
QY 2029 AGGGAAGTCCGAGCGCTTTTGGCAGAGATCAGGAGATGGTGGCAGAGTCCACGCTAGG 2088
DB 29815 GACGTGCGCGCGCGCGGACCGCTCGGCACCATCAGCAGGCGCGGACACCGGAAAG 29874
QY 2089 ATGATCAAGA 2098
DB 29875 CTCGTCTGA 29884
RESULT 15
US-09-010-928B-3/c
Sequence 3, Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2824 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: -
LOCATION: 1..2824
OTHER INFORMATION: /note="Flagelliform DNA sequence
OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722.
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2724
US-09-010-928B-3

Query Match 1.5%; Score 36.4; DB 2; Length 2824;
Best Local Similarity 48.5%; Pred. No. 2.5;
Matches 100; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 1922 ACCTACCGCTCAACACCTTCTCTCTCGGCGGCTGCGGCTCATGCTACACGAGCAG 1981
DB 678 ACCTCTCTCGGCTGCTCTCTCTCCAGCGCCACCAACGAGGTCAGCTCCACGAGG 619
QY 1982 TCCTGGAGCGGCTGTGCAAAACACCAACCCCGCGGAGCCAGCCGAGCTTCAGGGACTGCCGAG 2041

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 77.6831 Seconds
(without alignments)
11655.134 Million cell updates/sec

Title: US-09-555-342b-1_COPY_151_2501
Perfect score: 2351
Sequence: 1 ccttcaggaactctgtc.....attgaggagcgaagacga 2351

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	762	32.4	3094	10	US-09-925-297-220
2	297	12.6	379	10	US-09-960-352-154
3	292	12.4	400	10	US-09-960-352-156
4	248.6	10.6	408	10	US-09-783-590-3620
5	237.4	10.1	500	10	US-09-783-590-3621
6	190.8	8.1	335	10	US-09-783-590-3600
7	168.8	7.2	356	10	US-09-783-590-3575
8	154.6	6.6	426	10	US-09-815-343-1490
9	144	6.1	3984	10	US-09-848-294-1
10	142.8	6.1	2872	10	US-09-906-779-3
11	142.8	6.1	4125	12	US-10-044-090-520
12	133.2	5.7	288	10	US-09-783-590-3542
13	106	4.5	1718	9	US-09-764-868-51
14	105.2	4.5	716	9	US-09-764-868-475
15	81.2	3.5	466	10	US-09-728-445-750
16	64.4	2.7	2686	9	US-09-764-868-48
17	64.4	2.7	3380	10	US-09-799-799-1
18	64	2.7	1091	9	US-09-764-868-316
19	57.6	2.5	408	10	US-09-960-352-6773

20	57.6	2.5	417	12	US-10-044-090-624	Sequence 624, App
21	57.4	2.4	3957	12	US-10-005-467-1	Sequence 1, Appli
22	56.8	2.4	4080	10	US-09-962-436-342	Sequence 342, App
23	50.2	2.1	512	10	US-09-960-253-27	Sequence 27, Appl
24	50.2	2.1	2330	10	US-09-960-253-156	Sequence 156, App
25	50.2	2.1	3044	10	US-09-880-107-3718	Sequence 3718, Ap
26	50.2	2.1	3047	10	US-09-864-864-329	Sequence 329, App
27	50.2	2.1	3115	10	US-09-925-299-123	Sequence 123, App
28	43.4	1.8	3786	10	US-09-815-242-7865	Sequence 7865, Ap
29	42.4	1.8	1730	9	US-09-860-670-78	Sequence 78, Appl
30	42.4	1.8	1730	9	US-09-764-868-52	Sequence 52, Appl
31	42.4	1.8	1730	9	US-09-764-868-476	Sequence 476, App
32	40.8	1.7	2172	10	US-09-815-242-4038	Sequence 4038, Ap
33	40.4	1.7	598	10	US-09-770-149-991	Sequence 991, App
34	40.4	1.7	768	9	US-09-938-842A-812	Sequence 812, App
35	39.6	1.7	4722	10	US-09-962-055-14	Sequence 14, Appl
36	39.6	1.7	4722	12	US-10-023-529-14	Sequence 14, Appl
37	39.6	1.7	4722	12	US-10-023-523-14	Sequence 14, Appl
38	39.4	1.7	420	10	US-09-960-352-9773	Sequence 9773, Ap
39	38.8	1.7	300	10	US-09-998-598-1581	Sequence 1581, Ap
40	38.6	1.6	414	10	US-09-954-456-1461	Sequence 1461, Ap
41	38.4	1.6	1685	9	US-09-954-531-133	Sequence 133, App
42	38.4	1.6	1685	9	US-09-954-531-354	Sequence 354, App
43	38.4	1.6	1685	10	US-09-962-436-261	Sequence 261, App
44	38.4	1.6	2296	10	US-09-822-849A-259	Sequence 259, App
45	37.8	1.6	420	10	US-09-960-352-9532	Sequence 9532, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-220
; Sequence 220, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 220
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-220

Query Match 32.4%; Score 762; DB 10; Length 3094;
Best Local Similarity 100.0%; Pred. No. 2.5e-211;
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1590 GTGGTTTCAGACAGTGTGAGCAAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT 1649
Db 62 GTGGTTTCAGACAGTGTGAGCAAGAGCGCCATGCCGGAAGCACTGAAAGTCTCAT 121
QY 1650 ATTCCCGAATTTGACCTTTCACAAATTTCTACTAATTTCTCAAGAAATTCAGCA 1709
Db 122 ATTCCCGAATTTGACCTTTCACAAATTTCTACTAATTTCTCAAGAAATTCAGCA 181
QY 1710 ACGACTTGCCCTGTGGGAAGCGCGCTCAAATGCCCAATCAGAGATTACCAAGAAATCGG 1769

```
|||||
Db 182 ACAGACTTGCCTTGGGAAGCGCGCTCAAAATGCCAAATCAGAGATTACCAAGAAATCGG 241
QY 1770 CGATGTCATGCTCAAGAAATTCAGGGCATGAAGCAGCTCGGGCTCACCCTGTGGGAAGCA 1829
Db 242 CGATGTCATGCTCAAGAAATTCAGGGCATGAAGCAGCTCGGGCTCACCCTGTGGGAAGCA 301
QY 1830 CAGCGAGGCGCTTGGAGCGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 1889
Db 302 CAGCGAGGCGCTTGGAGCGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 361
QY 1890 CTCGAGAGACTTTGAGCTCGAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGG 1949
Db 362 CTCGAGAGACTTTGAGCTCGAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCTCGG 421
QY 1950 GCCACTGCAACCGCTCATGCACTACAGCAGAGTCTGGAGCGGCTGTGCAAAACACCAACC 2009
Db 422 GCCACTGCAACCGGCTCATGCACTACAGCAGAGTCTGGAGCGGCTGTGCAAAACACCAACC 481
QY 2010 GCGAGCGCAGCGGACTTCAGGGACTGCGAGCGGCTTTGGCAGAGATCACGGAGATGCT 2069
Db 482 GCGAGCGCAGCGGACTTCAGGGACTGCGAGCGGCTTTGGCAGAGATCACGGAGATGCT 541
QY 2070 GGCACAGCTCCAGCGTACGATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAACCTAA 2129
Db 542 GGCACAGCTCCAGCGTACGATGATCAAGATGAGAAATTTCCAGAAGCTGCACGAACCTAA 601
QY 2130 GAAAGATTGTATGGCATTGACAAATCTGTGTTCCGGGAAGGAGTTCATCGTCTGG 2189
Db 602 GAAAGATTGTATGGCATTGACAAATCTGTGTTCCGGGAAGGAGTTCATCGTCTGG 661
QY 2190 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTCTGTTCACGA 2249
Db 662 CAGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTCTGTTCACGA 721
QY 2250 CGTCTGCTATACAGAGCGGGGGCTGACGGCTCCATCATGTTTAAAGTTCACGGGCA 2309
Db 722 CGTCTGCTATACAGAGCGGGGGCTGACGGCTCCATCATGTTTAAAGTTCACGGGCA 781
QY 2310 GCTCCCGCTCTATGGCATGACGATTTCAGAGAGCGGAAGACGA 2351
Db 782 GCTCCCGCTCTATGGCATGACGATTTCAGAGAGCGGAAGACGA 823

RESULT 2
US-09-960-352-154
; Sequence 154, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 154
; LENGTH: 379
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E1-A9
US-09-960-352-154
Query Match 12.6%; Score 297; DB 10; Length 379;
Best Local Similarity 89.7%; Pred. No. 1.5e-76;
Matches 331; Conservative 0; Mismatches 35; Indels 3; Gaps 1;

QY 1485 TGAGGCGCGGAGGAAGAGATTCCCAACTGATAAAGGTACTTCATAGCTAAAGAACTGTC 1544
|||
Db 11 TGGGGCGCGGAGAAAGAGATTCCCAACTGACAAAGCATACTTCATTGCTTAAAGAAGTCTC 68
QY 1545 TACCACCGAGCGAACAATATCTGAAGGATCTCGAAGTATATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 69 CACCACTGAGAGACATATCTGAAGGATCTTGAAGTTCATCACTTCGTGTTTCAGAGCAC 128
|||||
QY 1605 AGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 129 AGTGAGCAAGAGGAGCTCCATGCCGGAAGCTTGAAGTCTCATATTTCCCGAAATTTGA 188
QY 1665 ACCTTTGCACAAATTTTCATACATAATTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTG 1724
|||||
Db 189 ACCTTTGCACAAATTTTCACACATATTTCTCAAGGACATTTGAGCAACGACTTGCCTGTG 248
|||||
QY 1725 GGAAGCGCGCTCAAAATGCCAAATCA--GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 249 GGAAGCGCGCTCGAATGCCACATCAGAGAGATTAACATAGAAATCGGAGATGTAGTCT 308
|||||
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTG-GCGGCTCACCTGTGGAAGCACAGCGAGCGCT 1840
|||||
```

```
QY 1545 TACCACCGAGCGAACAATATCTGAAGGATCTCGAAGTATATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 71 CACCACTGAGAGACATATCTGAAGGATCTTGAAGTATATCACTTCGTGTTTCAGAGCAC 130
QY 1605 AGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 131 AGTGAGCAAGAGGAGCTCCATGCCGGAAGCTTGAAGAGTCTCATATTTCCCGAAATTTGA 190
QY 1665 ACCTTTGCACAAATTTTCATACATAATTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTG 1724
|||||
Db 191 ACCTTTGCACAAATTTTCACAAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTG 250
QY 1725 GGAAGCGCGCTCAAAATGCCAAATCA--GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 251 GGAAGCGCGCTCGAATGCCACATCAGAGAGATTAACAGAGAAATCGGAGATGTATGCT 310
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTGCGGCTCACCTGTGGAAGCACAGCGAGCGCTT 1841
|||||
Db 311 GAAGAAGATTTCAGGGATGAAGCAACTGCGGCCCACTTGTGGAAGCACAGCGAGCGCT 370
QY 1842 GGAGGCGCT 1850
|||||
Db 371 GGAGGCGCT 379

RESULT 3
US-09-960-352-156
; Sequence 156, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 156
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E2-A9
US-09-960-352-156
Query Match 12.4%; Score 292; DB 10; Length 400;
Best Local Similarity 87.5%; Pred. No. 4.3e-75;
Matches 343; Conservative 0; Mismatches 45; Indels 4; Gaps 2;

QY 1485 TGAGGCGCGGAGGAAGAGATTCCCAACTGATAAAGGTACTTCATAGCTAAAGAACTGTC 1544
|||
Db 9 TGGGGCGCGGAGAAAGAGATTCCCAACTGACAAAGCATACTTCATTGCTTAAAGAAGTCTC 68
QY 1545 TACCACCGAGCGAACAATATCTGAAGGATCTCGAAGTATATCACTTCGTGTTTCAGAGCAC 1604
|||||
Db 69 CACCACTGAGAGACATATCTGAAGGATCTTGAAGTTCATCACTTCGTGTTTCAGAGCAC 128
|||||
QY 1605 AGTGAGCAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAAATTTGA 1664
|||||
Db 129 AGTGAGCAAGAGGAGCTCCATGCCGGAAGCTTGAAGTCTCATATTTCCCGAAATTTGA 188
QY 1665 ACCTTTGCACAAATTTTCATACATAATTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTG 1724
|||||
Db 189 ACCTTTGCACAAATTTTCACACATATTTCTCAAGGACATTTGAGCAACGACTTGCCTGTG 248
|||||
QY 1725 GGAAGCGCGCTCAAAATGCCAAATCA--GAGATTACCAAGAAATCGGCGATGTCATGCT 1781
|||||
Db 249 GGAAGCGCGCTCGAATGCCACATCAGAGAGATTAACATAGAAATCGGAGATGTAGTCT 308
|||||
QY 1782 GAAGAACATTTCAGGCGCATGAAGCACCTG-GCGGCTCACCTGTGGAAGCACAGCGAGCGCT 1840
|||||
```



```
Db 309 GAAGAACATTACGGGATGAAGCAACTGAGCTGCTCATTGTGGAAGCACAGCTAGGCC 368
QY 1841 TGGAGGCCCTGGAGAAATGGAATCAAGAGCTCC 1872
Db 369 TGGAGGCCCTGGACATCGGCATCCTGGGCCCC 400

RESULT 4
US-09-783-590-3620
; Sequence 3620, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3620
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (352)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (406)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3620

Query Match 10.6%; Score 248.6; DB 10; Length 408;
Best Local Similarity 85.0%; Pred. No. 1.9e-62;
Matches 347; Conservative 0; Mismatches 46; Indels 15; Gaps 6;

QY 1621 GCCATGCCGAAGCACTGAAAAGCTCATATCCCGAATTTTCAACCTTTGCACAAATTT 1680
Db 1 GGAGAGCGAGCACTGAAAAGCTCATATCCCGAATTTTCAACCTTTGCACAAATTT 60

QY 1681 CATACAAATTTTCTCAAGAAATTTGACACGACTTGCCCTGTGGGAAGCCGCTCAAAAT 1740
Db 61 CATACTAATTTTTCAGGAAATTTGACACGACTTGCCCTGTGGGAAGCCGCTCAAAAT 120

QY 1741 GCCCAATCAGAGATTACCAAGAATCGGCGATGTGCTGCTGAAGAACATTCAGGCGATG 1800
```

```
Db 121 GCCCAATCAGAGATTACCAAGAATCGGCGATGTGCTGCTGAAGAACATTCAGGCGATG 180
QY 1801 AAGCACTGCGGCTCACCTGTGTGAAGCACAGCGA-GGCCCTTGAGGCGCCTG---GAGAA 1856
Db 181 AAGCACTGCGGCTCACCTGTGTGAAGCACAGCGAGGCCCTTGAGGCGCCTGGGAGGAAT 240
QY 1857 TGGAAATCAAGAGCTCCCGGGCGGTGGAGA---ACTTCTGCAGAGACTTTTGTAGCTGCAGA 1912
Db 241 GGCAATCAAGAGCTCCCGGGCGGTGGAGA---ACTTCTGCAGAGACTTTTGTAGCTGCAGA 300
QY 1913 AGGTGTGTTTACCTTACCGCTCAACACCTTCTCTCTGCG-----GGCCACTGCACCGGCTCAT 1967
Db 301 AGGTGTGTTTACCTTACCGCTTCAACAACTTCTCTCTGNGGGCCACTGCGANGGNTCAT 360
QY 1968 -GCACTACAAGCAGGTCTCTGGAGC-GGCTGTGCAAAACACCAACCCGCG 2013
Db 361 GGCAATCAAGCAAGTTCTGGAGCGGTTGTGCAAAACANCCNGCG 408

RESULT 5
US-09-783-590-3621
; Sequence 3621, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3621
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (146)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (240)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (336)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (377)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (380)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (388)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (390)
```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (436)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (450)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (459)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (461)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (475)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (499)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3621

Query Match 10.1%; Score 237.4; DB 10; Length 500;
Best Local Similarity 83.3%; Pred. No. 3.9e-59;
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13;
QY 1621 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGGCACAATTT 1680
DB 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGGCACAATTT 60
QY 1681 CATACATAATTTTCAAGAAATTTAGCAACGACTTGCCTGTG-GGGAAGCGCGTCAAA 1739
DB 61 CATACATAATTTTCAAGAAATTTAGCAACGACTTGCCTGTG-GGGAAGCGCGTCAAA 120
QY 1740 TGCCCAATCAGAGATTACCAAGATCGCGATGTCATGCTGAAGGAACATTCAGGGCA 1798
DB 121 TGCCCAATCAGAGATTACCAAGATCGCGATGTCATGCTGAAGGAACATTCAGGGCA 180
QY 1799 TGAAGCACCTGGCGG-CTCACCTGT-GGAAGCACAGCG-AGGCCTTGGAGGCCCTGGAGA 1855
DB 181 TGAAGCACCTGGCGGTCACCTGTGGGAAGCACAGCGAGGCCCTTGGAGGCCCTGGAGN 240
QY 1856 A--TGAATCAAGAGTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 1910
DB 241 AATGGNATTCAGAGCTCCCGCGGCTGGAGGAGNCTTCTGCAGAGACTTTGGAGCTGCA 300
QY 1911 GAGGTGTG-TTACCTACCGCTCAACACCTT--CTTCCTGGCGGCACTGCACCGGCTCAT 1967
DB 301 GAAGGTGTGTTTACCTACCGCTCAACACCTTCTTCCTNCTGCGGCGCACTGCACCGGTTCT 360
QY 1968 GCACTACAAG--CAGGTCTCGAGGGCTGTGCAACACACCCCGCGGAGCCGCGCA- 2024
DB 361 GCACTACAGGCGAGGTTCTNCTGAGCGGNTTGTCCAAACAAACCCGATCGNGGCGAAG 420
QY 2025 -----CTTCAGGAGTCCGAGCGCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGT 2078
DB 421 GCCGATTTTCAGGAGTCCGAGCGCG-TTTNGCAGNGTTNANGAGTTCTCTGGGANANTT 479
QY 2079 CCACGGTACGATCAAA 2096
DB 480 CCAGGTTGGTTGTTCAA 497

RESULT 6

US-09-783-590-3600
; Sequence 3600, Application US/097833590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Haseltine, Patrick J.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3600
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (29)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (36)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (157)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (184)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (254)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (271)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (333)

QY 795 GGATTTCTGCAAGTCTTCTGGAAATCTGTGTGAAATCATCATGCTTCTTT 846
Db 896 CGATCTTGCAAAACTGTGGAAATCCCTGTGTGAGCAACCATACGTTCTTT 947

RESULT 10
US-09-906-779-3
; Sequence 3, Application US/09906779
; Patent No. US2002006484A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: Human Protein Tyrosine Phosphatase Polynucleotides, Polypeptides,
; TITLE OF INVENTION: Antibodies
; FILE REFERENCE: PT040PI
; CURRENT APPLICATION NUMBER: US/09/906,779
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US01/01563
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/176,306
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2872
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-906-779-3

Query Match 6.1%; Score 142.8; DB 10; Length 2872;
Best Local Similarity 50.3%; Pred. No. 4.2e-31;
Matches 463; Conservative 0; Mismatches 442; Indels 15; Gaps 4;

QY 35 TGCTGATGACACCCAGGAGCAATTTGAAGTTCCACAAGAGCTCTCTGGGAAGTGTCTGC 94
Db 54 TGCTCGACGGGACCGAAGTGAGGTGAGCTGCCGAACATGCCAAGGCCAGGATTTGT 113
QY 95 TGGATCAGTTTCCACACACCTCAACCTCGTGGGAAGGTGACTATTTTGGCCTCGAGTTTC 154
Db 114 TTGATCAGATTGTGTACCACTTGAGCTGTGGAAACAGATTACTTTGGCCTCCAGTTCC 173
QY 155 CTGATCACAAAAGATCACGGTGTGCTGTGATCTCTTAACCCATGTGCAACAGATTA 214
Db 174 TCGACTCTGCCAGGTGGCGACTGCTGTGATCTGCCAACCCTAATAAAGCAGATGA 233
QY 215 GAAGGCCAAGACAGCTGTGTGTTGAAGTTTGTGTGAATTTCTTCGCGCTGACCAACAC 274
Db 234 AAATTGGACCTGCTTATGCTTTACACTTTCGAGTTAAATACTATTCTTCAGAACCAACA 293
QY 275 AACTCCAAGNAGAACTCACAGTACTGTTCGCGCTGAGGTGAAGAGGACTTGGCTC 334
Db 294 ACCTTCGTGAGGAGTTTACAAGGTACTGTGTTGTTTACAACTCAGGCATGACATTCTTT 353
QY 335 AAGCAGGTGTGAGCTGTGTAATGACACAGCGCAGCTCTCTTGATTTTACACATTTGTCGAAT 394
Db 354 CTGGAATTTGAATGCCCTTATGAACAGCTGTGGAATTAGCTGTCTGTCTGCTACAAG 413
QY 395 CTGAGATTGGGGATTTGATGAAGCCTTGGAACA---GAGACACTTTAGCAAAAATAAAT 451
Db 414 CGGAGCTTGGGGAGTGGAGCTTCCAGAACACACACACAGAGCTGTGTGCTGAGTTTCGGT 473
QY 452 ACATACC---TCAGCAGAGCGCACTAGAGGACAAAATCGTGGAAATTCACCATAAACACA 508
Db 474 TCATTCGAATACAGACAGCAAGTGAATTTGATATCTTCCAGAGATGGAAGAGTGCA 533
QY 509 TTGGCAAAACACCCAGAGATCAGATTTCCAGCTCTAGAGATTGCCGTGGCTAGAGA 568
Db 534 GGGGAAGAGCCCTGCCAGCGGCACTCTCTCTATCTGAATAAGCGAAGTGGCTGGAA 593
QY 569 TGTATGGAATCCGGTTTCCCGCCGCGCAAGGACAGGGAAGCAAGATCAATCTGGCCG 628
Db 594 TGTATGGGTAGACATGCAGTTTGCAGGGGAAGAGATGCTGTGATATTCTCTTTGGAC 653
QY 629 TTGCCAACAGGGGAATTTAGTGTTTTTCAGGGTTTCACTAAGATCAATGCCCTCAACTGGG 688

Db 654 TGACCCCGCAGGAGCATATTAAATCTTTGAAGGAGCTAACAAAATAGGCTTATTTTGGC 713
QY 689 CCAAGGTGCGGAAGCTTGAGCTTCAAGAGAAAGCGCTTTCTCATCAAGCTCCCGCCAGATG 748
Db 714 CTAATAATTACCAAAATGGATTTTAAAAGAGCAAAATGACACTCGTGGTGCAGAGATG 773
QY 749 CCAAT-----AGTGCGTACAGGATACCTTGGAAATTCCTGTGATGGCCAGTGGGATTTCT 802
Db 774 ATGATCAGGGAGCTGAGCAAGAGACACACGTTTGTGTTCCGGTTAGACAGTCCAGGACCT 833
QY 803 GCAAGTCCCTTCGGAATACTGTGTGACATCATGCTTCTTTAGACATTTTGAAGAGC 862
Db 834 GCAACACACCTTTGGAAAGTGTGCAGTTTGACACACACCAATCTTCCGACTGCGGACGCCAG 893
QY 863 CCAAAACCAAGCCCAAGC---CCGTCTCTTTAGCCGGGTGCATCTTTCGGTTTCAGTG 919
Db 894 GAAACAGCAAAATCCAATAGATCCGACTTTATCAGGCTGGGCTCTCGCTTCAGATTCAGTG 953
QY 920 GTCGGACTCAGAAAGCAGGTT 939
Db 954 GGGGACAGAATATCAAGCT 973

RESULT 11
US-10-044-090-620
; Sequence 620, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 620
; LENGTH: 4125
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 347941.18
; NAME/KEY: unsure
; LOCATION: 3891, 3894, 3896-3897, 3905, 3908, 3915, 4019, 4035-4037, 4067, 4083
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-620

Query Match 6.1%; Score 142.8; DB 12; Length 4125;
Best Local Similarity 51.3%; Pred. No. 5.4e-31;
Matches 435; Conservative 0; Mismatches 402; Indels 11; Gaps 4;

QY 10 ARACTCGTGTCCATCAAAATCCAGATGCTGGATGCAAGTTTGAACCCACCTCAACCTCGTGGAA 69
Db 515 AAAGCATGCAAGTCAAAAGTGATCTCTCGATGATCAGATATACCTGTGATGAGAG 574
QY 70 CAAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAAGTTTGAACCCACCTCAACCTCGTGGAA 129
Db 575 ARAAGCTCCAGGACAGAGTGTGTTGATAAAGTGTGTGAACACTTGAACCTTGTAGAG 634
QY 130 GGTGACTATTGGCCCTCGAGTTTCTGTATCACAAAAGATCAGGTTGGGTGGATCTC 189
Db 635 AAAGACTACTTTGGGCTTACGTATCGAGATGCTGAAACACGAGAAGATTGGTTGGACCT 694
QY 190 CTAARACCCATCTGAAACAGATTAGAAGCCCAACACAGCTTGTGTTAAGTTTGGTG 249
Db 695 GCTAAGGAATAAAAAACAGGTTCCGAAGTGGTGTGGCACTTTTCA---TTTAATGTG 750
QY 250 AAATCTTTCCGCTCAGCCACACAACTCCAAAGAAAGAACTCAAGAGTACCTGTTCGG 309
Db 751 AAATTTATCCACCAGCCCTGCCCAATATCT-GAAGATATACACAGTACTACCTCTGC 809
QY 310 CTGCAAGTGAAGCAGGACTTTGGCTCAAGGCAAGGTTGACCTGTAAATGACACAGCCAGCT 369

Db 810 TTGCAGTTGGAGATGACATCGTGTCCGGAAGGCTGCCCTGCTCCTTTGTTACCCCTGGCC 869
QY 370 CTTCTGATTTTCACACATTTGCAATCTGAGATTTGGGATTTTGA----TGAAGCCCTTGGAC 426
Db 870 TTGCTGGGCTCCTCATGCTTCCAGTTCAGAGCTCGGAGACTATGACCCAGATGAATGTGGG 929
QY 427 AGAGACACTTAGCAAAAAATAAATACATACCTCAGCAAG---ACGCACCTAGAGGACAAA 483
Db 930 AGCGATTACATTAGTAGTGAGTTCCGCTTTGACCAACACACTAAAGAACTGGAAGACAAA 989
QY 484 ATCGTGGAAATTTACCATTAACACATTTGGCAAAACACAGCAGAAATFCAGATTTCCAGCTC 543
Db 990 GTGATCGAGCTGCACAAGAGCCACAGAGAAATGACGCCAGCAGAGCAGAGATGCATTTTC 1049
QY 544 CTAGAGATTTGCCGCTGGCTAGAGATGTATGGAATCCGGTTGACCCGCCAAGGACAG 603
Db 1050 TTGGAANAATGCCAAAAAATAATATCAATGTATGGGTAGATTTTACATCATGCTAAGGACTCA 1109
QY 604 GAAGGCACGAAGATCAATCTGGCGGTGCCAACACACGGGAATTTCTAGTGTTCAGGGTTTC 663
Db 1110 GAAGGGGTAGAAATATGTAGAGTTTGTGCAAGTGGTCTGTGATATATCCGACCCG 1169
QY 664 ACTAAGATCAATGCCCTCAACTGGGCGCAAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGC 723
Db 1170 CTGCGAATAAACAGATTTGCCTGGCCCAAGTTCTAAAGATTTTACATAAACCGGAACAAC 1229
QY 724 TTTCTCATCAAGCTCGGCGGCAGATGCCAATAGTGGTACAGGATACCTTGAATTCCTG 783
Db 1230 TTTTACATTAAGATCGGCGCGGAGAGTTTGAACAAATTTGAAAGCACCATTTGGGTTTAAG 1289
QY 784 ATGGCAGTCTGGGATTTCTGCAAGTCTTCTGAAAATCTGTTGAACATCATGCTTTC 843
Db 1290 CTGCCAAACCATCGAGCTGCCAAGCGTTTATGGAAGTATGTGTGAGATCATATACATTT 1349
QY 844 TTTAGACT 851
Db 1350 TTCAGACT 1357

RESULT 12
US-09-783-590-3542
; Sequence 3542, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3542
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (72)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (75)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (93)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (129)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (142)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (155)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (156)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (166)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (182)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (253)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (263)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (270)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (282)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3542
Query Match 5.7%; Score 133.2; DB 10; Length 288;
Best Local Similarity 85.8%; Pred. No. 5.8e-29;
Matches 188; Conservative 0; Mismatches 27; Indels 4; Gaps 4;
QY 1628 CGGAAGCACTGAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAAATTTCTACTACTA 1687
Db 8 CGGAAGCACTGAAAGTCTCAAAATTTCCCAATTTTNAACCTTTTGCAAAATTTTAAACTA 67
QY 1688 ATTTTCTCAAGAAATTTGACCAACGACTTGCCTGTGGGAAGCCGCTCAAAATGCCCAAA 1747
Db 68 ATTTTNAAGAAATTTAAGCAACGNCCTTGCCTGTGGAAAGCCGCTCAAAATGCCCAAA 127
QY 1748 TCAGAGATTACCAAGAAATCGCGGATGT-CATGCTGAAG-AACATTCAGGGGATGAAGCA 1805
Db 128 TNA-AGATTTACCAANAATCGCGGATGTNNATGCTGAAGNANCATTCAGGGGATNGGCA 186
QY 1806 CTGGCGGC-TCACCTGTGGAAGCACAGCGAGCCCTTG 1843
Db 187 CTGGCGGCTTCACCTGTGGGAAGCACGCGAGGCGCTTG 225
RESULT 13

US-09-764-868-51
; Sequence 51, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR APPLICATION DATA: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-51

Query Match 4.5%; Score 106; DB 9; Length 1718;
Best Local Similarity 64.9%; Pred. No. 1.6e-20;
Matches 157; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 2110 CAGAACTCAGCAACTCAGAAAGATTGATGGCATTGACAATCTTGGTTCGGGA 2169
DB 13 CAGAACTCAGCGAGCTGCAGCGGGACCTGGTGGCATAGAGAACCTCATGCTCTGCG 72
QY 2170 AGGAGTTCATCGTCTGGGCAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGC 2229
DB 73 AGGAGTTCATCGTGCAGGCTGCCTTCACAGCTCACCAGAAAGGCTTCAGCAGAGG 132
QY 2230 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCCGGGGCTGACGGCTCCAAT 2289
DB 133 ATGTTCTTCTGTTCTCAGATATGTTGCTGTACACAAGCAAGAGGAGTTCGAGGACCGC 192
QY 2290 CAGTTTAAAGTCCAGCGAGCTCCGCTCTATGGCATGACGATTGAGAGAGCGAAGAC 2349
DB 193 CACTTCGGATCCGGGGCTCTCTCCCTCCCAAGGCATCTGTTGGAAGAAAGTGATAAC 252
QY 2350 GA 2351
DB 253 GA 254

RESULT 14
US-09-764-868-475
; Sequence 475, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 475
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (715)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-475

Query Match 4.5%; Score 105.2; DB 9; Length 716;
Best Local Similarity 64.0%; Pred. No. 1.5e-20;
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;
QY 2110 CAGAACTCAGCAACTCAGAAAGATTGATGGCATTGACAATCTTGGTTCGGGA 2169
DB 3 CAGAACTCAGCGAGCTGCWKCGGACCTGGTGGCATAGAGAACCTCATGCTCTGCGC 62

QY 2170 AGGAGTTCATCGTCTGGGCAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGC 2229
DB 63 AGGAGTTCATCGTGCAGGCTGCCTTCACAGCTCACCAGAAAGGCTTCAGCAGAGG 122
QY 2230 ATGTTCTTCTGTTCAACGAGCTCTGCTGTATACAGAGCCGGGGCTGACGGCTCCAAT 2289
DB 123 ATGTTCTTCTGTTCTCAGATATGTTGCTGTACACAAGCAAGAGGAGTTCGAGGACCGC 182
QY 2290 CAGTTTAAAGTCCAGCGGAGCTCCGCTCTATGGCATGACGATTGAGAGAGCGAAGAC 2349
DB 183 CACTTCGGATCCGGGGCTCTCTCCCTCCCAAGGCATCTGTTGGAAGAAAGTGATAAC 242
QY 2350 GA 2351
DB 243 GA 244

RESULT 15
US-09-728-445-750
; Sequence 750, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; TITLE OF INVENTION: Animals
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 750
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(466)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-750

Query Match 3.5%; Score 81.2; DB 10; Length 466;
Best Local Similarity 58.3%; Pred. No. 1.1e-13;
Matches 162; Conservative 0; Mismatches 113; Indels 3; Gaps 1;
QY 24 CAAATCCAGATCGTGGATGACACCCAGGAGCATTTGAAGTTCCACAAAGAGCTCCTGG 83
DB 118 CAACATCTTCTTCGTGATAACACTGTACAGGCTTCAGAGTTTACAAACATGATCAGG 177
QY 84 GAAGTGTGCTGGATGACAGTTTGCACACCTCAACCTCGTGGAGAGTGACTATTTGG 143
DB 178 GCAAGTTCGTGGATATAGTCTTCAAGCATCTTGATTTGACTGAGCGAGACTATTTGG 237
QY 144 CCTCGAGTTTCTCT---GATCACAAGAGATCAGGCTGCTGGATCTCTCTAAACCCAT 200
DB 238 TTTACAGTTGGCTGACGATTCCACAGATAACCCAAAGGTGGCTGGATCCAAACCAACCAAT 297
QY 201 TGTGAACACAGATTAGAAGGCCAACGACCGTTGTTGTTAAGTTTGTGGTCAAAATCTTTCC 260
DB 298 AAGGAAGCAGCTAAGAGAGAGGATCACCCTTACAAATTTGAATTTAGAGTCAAAATCTTTGT 357
QY 261 GCTTGACACACACACTCCAAAGAAAGAACTCAAGGT 298
DB 358 AAGTGACCCCAACAAGTTTACAAGAGAGATACAAGGT 395

Search completed: December 3, 2002, 22:26:06
Job time : 115.683 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 3, 2002, 16:31:10 ; Search time 2531.22 Seconds
(without alignments)
15042.359 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_2501
Perfect score: 2351
Sequence: 1 ccttcagagaaactcgtgc.....attgagagagcgaagacga 2351

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_hic:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_hic:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_mam:**
- 24: em_gss_mus:**
- 25: em_gss_other:**
- 26: em_gss_pro:**
- 27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	883.4	37.6	912	14	BQ881649
2	812.6	34.6	926	14	BQ706499
3	781.8	33.3	896	14	BQ672632
4	781	33.2	923	14	BQ950768
5	762	32.4	774	13	B1760610
6	761.2	32.4	804	12	BG747792

7	757.2	32.2	865	10	BE260677
8	755.8	32.1	784	12	BG763918
9	744.8	31.7	1043	14	BQ072025
10	726.2	30.9	901	12	BG475554
11	705.4	30.0	898	12	BE745887
12	704.6	30.0	835	12	BG829192
13	700.2	29.8	902	12	BE910036
14	687.6	29.2	721	9	AU132546
15	682	29.0	988	14	BQ674703
16	679.6	28.9	942	12	BG323704
17	678.2	28.8	891	12	BG767698
18	660	28.1	680	12	BG825738
19	628.4	26.7	909	14	BQ646351
20	608.4	25.9	968	12	BF206873
21	607.6	25.8	934	12	BF314265
22	605.6	25.8	692	12	BG116225
23	598.8	25.5	704	12	BG769615
24	592	25.2	621	12	BG770181
25	591	25.1	861	12	BG764061
26	590.4	25.1	932	12	BF686586
27	577.6	24.6	890	12	BE907778
28	560.6	23.8	788	12	BG750463
29	551.8	23.5	954	12	BF304259
30	551	23.4	560	10	AW411445
31	544	23.1	618	14	BQ331885
32	529.8	22.5	627	10	AV729533
33	526	22.4	859	9	AL537488
34	507	21.6	649	14	BQ745892
35	494.4	21.0	496	14	BM705217
36	491.2	20.9	995	13	BM549907
37	478.4	20.3	939	12	BF316537
38	475.4	20.2	489	9	AL121548
39	468	19.9	736	14	BM963756
40	463	19.7	861	12	BG420356
41	457	19.4	1026	13	BM551766
42	456.4	19.4	1002	12	BG323892
43	454	19.3	454	14	BM728340
44	446.2	19.0	961	14	BQ068709
45	445.4	18.9	708	13	BI696633

ALIGNMENTS

RESULT 1
BQ881649
LOCUS BQ881649 912 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT_8728249 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6339275
5', mRNA sequence.
ACCESSION BQ881649
VERSION BQ881649.1 GI:22273657
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 912)
AUTHORS NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>,
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2533 row: n column: 12
High quality sequence stop: 684.
Location/Qualifiers 1..912

FEATURES
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6339275"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter:
GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      247 a   235 c   252 g   178 t
ORIGIN
Query Match      37.6%; Score 883.4; DB 14; Length 912;
Best Local Similarity 98.7%; Pred. No. 6.e-235;
Matches 901; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 442 AAAAAATAATACATACCTCAGCAAGACGACTAGAGGACAAAATCGTGAATTTCCACCAT 501
Db 1 AAAAAATAATACATACCTCAGCAAGACGACTAGAGGACAAAATCGTGAATTTCCACCAT 60
QY 502 AACCACATTGGCAACACACAGCAATCAGATTTCAGCTCCTAGAGATTGCCGCTCGG 561
Db 61 AACCATTTGGACAACACACAGCAATCAGATTTCAGCTCCTAGAGATTGCCGCTCGG 120
QY 562 CTAGAGATGTATGGAATCCGGTTGCACCCGCCAAGGACAGGAGGACGACGAAGATCAAT 621
Db 121 CTAGAGATGTATGGAATCCGGTTGCACCCGCCAAGGACAGGAGGACGACGAAGATCAAT 180
QY 622 CTGGCGGTTCGCAACACGCGGAATTCAGTGTTCAGGGTTTCACCTAAGATCAATGCCCTTC 681
Db 181 CTGGCGGTTCGCAACACGCGGAATTCAGTGTTCAGGGTTTCACCTAAGATCAATGCCCTTC 240
QY 682 AACTGGCCCAAGTCGCGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGG 741
Db 241 AACTGGCCCAAGTCGCGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGG 300
QY 742 CCAGATGCCAATAGTCGCTPACAGGATACCTTTGGAAATTCCTGATGCCAGTCGGGATTC 801
Db 301 CCAGATGCCAATAGTCGCTPACAGGATACCTTTGGAAATTCCTGATGCCAGTCGGGATTC 360
QY 802 TCGAAGTCCTTCTGGAATACTGTGTGACATCATGCTTCTTTAGACTTTTTCGAAGAG 861
Db 361 TCGAAGTCCTTCTGGAATACTGTGTGACATCATGCTTCTTTAGACTTTTTCGAAGAG 420
QY 862 CCACAAACCAAGCCCAAGCCGCTCTCTTTAGCCGGGTCTATCATTTTCGGTTTCAGTGGT 921
Db 421 CCACAAACCAAGCCCAAGCCGCTCTCTTTAGCCGGGTCTATCATTTTCGGTTTCAGTGGT 480
QY 922 CGGACTCTAGAAGCAGGTTCCTGACTATGTTTAAAGAGGAGGACATAGAAGGTGCAGTTT 981
Db 481 CGGACTCTAGAAGCAGGTTCCTGACTATGTTTAAAGAGGAGGACATAGAAGGTGCAGTTT 540
QY 982 GAAAGGAAGCAGCAAGATTCATTCATCCGGAGCCCTTCTTCACAGCCTACAACTG 1041
Db 541 GAAAGGAAGCAGCAAGATTCATTCATCCGGAGCCCTTCTTCACAGCCTACAACTG 600
QY 1042 AATTTCGGAAGTCTCGAGCAGTCTCAGCAGACGACGACGCTTACATTTTGGAGAAGGTGCC 1101
Db 601 AATTTCGGAAGTCTCGAGCAGTCTCAGCAGACGACGACGCTTACATTTTGGAGAAGGTGCC 560
QY 1102 GAATCTCCAGGGGGCCAGAGCTTCGCGGCGAGGAAAGGAAACCGAAGGTTTTCGCCCGGGGAG 1161
Db 661 GAATCTCCAGGGGGCCAGAGCTTCGCGGCGAGGAAAGGAAACCGAAGGTTTTCGCCCGGGGAG 720
QY 1162 CCGGGGTCCGACCCCGAGCCCTCGCGGAGGAGAACGCCCGCGGGTTAACAGCAGCGGCAC 1221
Db 721 CCGGGGTCCGACCCCGAGCCCTCGCGGAGGAGAACGCCCGCGGGTTAACAGCAGCGGCAC 780
```

```
QY 1222 GGAGCGCCTCGCGCCGCCACGAGGAGAGGAGGAGGTCGTTAAGGATPAGGACCCAGCAG 1281
Db 781 GGAGCGCCTTCGCGCGCCGCCACGAGGAGGAGGAGGTCGTTAAGGATPAGGACCCAGCAG 840
QY 1282 AGTAAACCTCAGCCGCCCGCCAGCAGCAGCTCCCTGACTGGCAGTCCTCACATTTCC 1341
Db 841 AGTAAACCTCAGCCGCCCGCCAGCAGCAGCTCCCTGACTGGCAGT-CTCACCTTTTC 899
QY 1342 GAGCTGCTGTGA 1354
Db 900 GAGCTGCTGTGA 912

RESULT 2
LOCUS      BQ706499      926 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION AGENCOURT_8474935 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301634
5', mRNA sequence.
ACCESSION  BQ706499
VERSION    BQ706499.1 GI:21845398
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 926)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. Mark Watson
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone Distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2517 row: n column: 03
          High quality sequence stop: 601.
          Location/Qualifiers
            1..926
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:6301634"
              /clone_lib="NIH_MGC_113"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGCAGGAG(G). Library constructed by Ling Hong in the
              laboratory of Gerald M. Rubin (University of California,
              Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
              Superscript II RT (Life Technologies). Note: this is a
              NIH_MGC Library."
BASE COUNT      242 a   247 c   254 g   183 t
ORIGIN
Query Match      34.6%; Score 812.6; DB 14; Length 926;
Best Local Similarity 98.8%; Pred. No. 3.9e-215;
Matches 829; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1470 GACGGACCATGAGGATGAGGCGCGGAGGAGATTCCTCAACTGATAAGCGTACTTCAT 1529
Db 1 GACGGACCATGAGGATGAGGCGCGGAGGAGATTCCTCAACTGATAAGCGTACTTCAT 60
QY 1530 AGCTAAGGAAGTGTCTACCACCGAGCGAACATATCTGAAGGATCTCGAAGTTATCATTTC 1589
Db 61 AGCTAAGGAAGTGTCTACCACCGAGCGAACATATCTGAAGGATCTCGAAGTTATCATTTC 120
QY 1590 GTGGTTTCAGACACAGTGCAGCAAGAGAGCCCATGCCGGGAGCAGCTGAAAGTCTCAT 1649
Db 1 GTGGTTTCAGACACAGTGCAGCAAGAGAGCCCATGCCGGGAGCAGCTGAAAGTCTCAT 1649
```



```

|||||
Db 721 ACCTTCTCTCTGGCGCACTGGACCGGCTCATGCATACAGCAAGCGCTGGAAGCGGT 780
|||||
QY 1995 GTCAAAACACCCCGCG-AGCCAGCGGACTTACGGGACTGCC 2038
|||||
Db 781 GTCAAAACACCCCGCGGAGCGCGCAACTTCAGGGAACGGC 825
|||||

RESULT 4
BO950768 923 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8842056 Lupski_sciatic_nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6204242 5', mRNA sequence.
ACCESSION BO950768
VERSION BO950768.1 GI:23266246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13625 row: d column: 03
High quality sequence stop: 585.
Location/Qualifiers
1..923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204242"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-GACTAGTTCTAGATCGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 238 a 257 c 239 g 189 t
ORIGIN
Query Match 33.2%; Score 781; DB 14; Length 923;
Best Local Similarity 96.0%; Pred. No. 2.6e-20;
Matches 834; Conservative 0; Mismatches 30; Indels 5; Gaps 3;

QY 1298 CGCAGCCAGCAGCGTCCCTGACTGGCGAGTCCTCAGCTTCCGAGCTGTCTGTGAAC 1357
|||||
Db 31 CTCATCCCCAACAGCTCCCTGACTGGCAGTCCCTCAGCTTCCGAGCTGTCTGTGAAC 90
|||||
QY 1358 CGCAGGGGAGTGCCCTGCGCAACGTGACCTGTCTCCCAACCTGAGCCCGCACCA 1417
|||||
Db 91 CGCAGGGGAGTGCCCTGCGCAACGTGACCTGTCTCCCAACCTGAGCCCGCACCA 150
|||||
QY 1418 ACAGGCTCTCCCTGATCAGCCCGCTGCTGAATGACCGAGCTGCCCGGAGCGAG 1477
|||||
Db 151 ACAGGCTCTCCCTGATCAGCCCGCTGCTGAATGACCGAGCTGCCCGGAGCGAG 210
|||||

```

```

QY 1478 ATGAGGATGAGGCGCGGAGGAAGAGATTCCCAACTGATAAAGCGTACTTCATAGCTAAG 1537
|||||
Db 211 ATGAGGATGAGGCGCGGAGGAAGAGATTCCCAACTGATAAAGCGTACTTCATAGCTAAG 270
|||||
QY 1538 AAGTGTCTACCAACCGAGGCAACATATCTGAAGGATCTCGAAGTTATCACTTGTGGTTTC 1597
|||||
Db 271 AAGTGTCTACCAACCGAGGCAACATATCTGAAGGATCTCGAAGTTATCACTTGTGGTTTC 330
|||||
QY 1598 AGAGCACAGTGAGCAAAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTCCCGA 1657
|||||
Db 331 AGAGCACAGTGAGCAAAAGAGGAGCGCATGCCGGAAGCACTGAAAAGTCTCATATTCCCGA 390
|||||
QY 1658 ATTTTGAACCTTTTGCACAAATTTTCACTAATTTTCTCAAGGAAATTGAGCAACGACTTG 1717
|||||
Db 391 ATTTTGAACCTTTTGCACAAATTTTCACTAATTTTCTCAAGGAAATTGAGCAACGACTTG 450
|||||
QY 1718 CCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGCA 1777
|||||
Db 451 CCTCTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGCA 510
|||||
QY 1778 TGCTGAAGAACAATTCAGGCGCATGAAGCACTGGCGGCTCACTGTGGAAGCACACGGAGG 1837
|||||
Db 511 TGCTGAAGAACAATTCAGGCGCATGAAGCACTGGCGGCTCACTGTGGAAGCACACGGAGG 570
|||||
QY 1838 CCTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGGCGCTGGAGAACTTCTGCAGAG 1897
|||||
Db 571 CCTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGGCGCTGGAGAACTTCTGCAGAG 630
|||||
QY 1898 ACTTTGAGCTGCGAAGGTGTGTACCTACCGCTCAACACCTTCTCTGCGGCGCACTGC 1957
|||||
Db 631 ACTTTGAGCTGCGAAGGTGTGTACCTACCGCTCAACACCTTCTCTGCGGCGCACTGC 690
|||||
QY 1958 ACCGCTCATGCACTACAAGAGGTCTTGAGCGGCTGTGCAAAACACACCGCGCG--AG 2015
|||||
Db 691 ACCGCTCATGCACTACAAGAGGTCTTGAGCGGCTGTGCAAAACACACCGCGCGGAGC 750
|||||
QY 2016 CCACGCCACTTCAGGAGCTGCCGAGCGC--TTTGGCAGAGATCAC--GGAGATGCTGCG 2072
|||||
Db 751 CACGCCCACTTCAGGAGCTGCCGAGCGCTTTGGCAGAGATCACCGGAAATGGTGGC 810
|||||
QY 2073 ACAGTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAAGTGCACGAACTCAAGAA 2132
|||||
Db 811 ACAGTCCACGGTACGATGATGATGAGAAATTTCCGGAAGCTGGCCCAACTCAGAAA 870
|||||
QY 2133 AGATTGATTGGCATTGACAACTTTGTGG 2161
|||||
Db 871 AGATTGATTGGCATTGACAACTTCTGGGG 899
|||||

```

```

RESULT 5
LOCUS BI760610
DEFINITION 603044980F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185411 5',
mRNA sequence.
ACCESSION BI760610
VERSION BI760610.1 GI:15752188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 774)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM11462 row: p column: 20
High quality sequence stop: 774.
Location/Qualifiers
1. .774

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
200 a 214 c 226 g 134 t

BASE COUNT

ORIGIN
Query Match 32.4%; Score 762; DB 13; Length 774;
Best Local Similarity 99.9%; Pred. No. 4.6e-201;
Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 958 GGAGGACATAAGAGTGCAGTTTGAAGGAAGCAGCAGCAAGATTTCATTCCTCGGAGC 1017
Db 1 GGAGGACATAAGAGTGCAGTTTGAAGGAAGCAGCAGCAAGATTTCATTCCTCGGAGC 60
QY 1018 CTTGCTTCACAGCCTCAGAACTGAAATCCGAGAGTGTCTGAGCAGTCTCAGCAGCACC 1077
Db 61 CTTGCTTCACAGCCTCAGAACTGAAATCCGAGAGTGTCTGAGCAGTCTCAGCAGCACC 120
QY 1078 AGCCTTTACA-TTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCCGGCGGAGGAAA 1136
Db 121 AGCCTTACAGTTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGTCCGGCGGAGGAAA 180
QY 1137 GGAACCGAGTTTCCGCGGGGAGCGGGGTGCGACCCGAGCCCTGCGCCGAGGAGAG 1196
Db 181 GGAACCGAGTTTCCGCGGGGAGCGGGGTGCGACCCGAGCCCTGCGCCGAGGAGAG 240
QY 1197 CCCCGGGGTAACAGCAGCGCAGCGAGCCGCTCGGCCCCACAGCAGGAAGAGGA 1256
Db 241 CCCCGGGGTAACAGCAGCGCAGCGAGCGGCTCGGCCCCACAGCAGGAAGAGGAGGA 300
QY 1257 GGTCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCCAGCAGCAGAGCTC 1316
Db 301 GGTCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCGCCAGCAGCAGAGCTC 360
QY 1317 CTTGACTGGCAGTCTCACCTTCCTGAGCTGTCTGTGAATCGAGGGGGAGTGGCCCC 1376
Db 361 CTTGACTGGCAGTCTCACCTTCCTGAGCTGTCTGTGAATCGAGGGGGAGTGGCCCC 420
QY 1377 TGCAACGTCAGCTGTCTCCCAACCTGAGCCCGCAGCAGCAGCAGCTCCCTTGAT 1436
Db 421 TGCAACGTCAGCTGTCTCCCAACCTGAGCCCGCAGCAGCAGCAGCTCCCTTGAT 480
QY 1437 CAGCCCGCTGCTGAATGACCCAGCGCTGCCCGCCGAGCGAGATGAGGATGAGGCGCGGAG 1496
Db 481 CAGCCCGCTGCTGAATGACCCAGCGCTGCCCGCCGAGCGAGATGAGGATGAGGCGCGGAG 540
QY 1497 GAGAGATTTCCCACTGATTAAGCGTACTTTCATAGCTTAAGGAAGTGTCTACCCAGCGC 1556
Db 541 GAGAGATTTCCCACTGATTAAGCGTACTTTCATAGCTTAAGGAAGTGTCTACCCAGCGC 600
QY 1557 AACATATCTGAAGSNTCTCGAAGTTTATCACTTCGTGGTTTCAGAGCAGCAGTGAGCAAGA 1616
Db 601 AACATATCTGAAGSNTCTCGAAGTTTATCACTTCGTGGTTTCAGAGCAGCAGTGAGCAAGA 660

QY 1617 GGAGCCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTGAACCTTGCACAA 1676
Db 661 GGAGCCATGCCGGAAGCACTGAAAGTCTCATATATCCCGAATTTGAACCTTGCACAA 720
QY 1677 ATTTCTACTAATTTCTCAAGGAATTTGAGCAAGCACTTGCCTCTGGGGAAG 1730
Db 721 ATTTCTACTAATTTCTCAAGGAATTTGAGCAAGCACTTGCCTCTGGGGAAG 774

RESULT 6
BG747792
LOCUS
DEFINITION
60270523F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4841975 5',
mRNA sequence.
ACCESSION
BG747792
VERSION
BG747792.1 GI:14058445
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 804)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1675 row: j column: 24
High quality sequence stop: 796.
Location/Qualifiers
1. .804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
225 a 191 c 209 g 179 t

BASE COUNT
ORIGIN
Query Match 32.4%; Score 761.2; DB 12; Length 804;
Best Local Similarity 99.1%; Pred. No. 7.9e-201;
Matches 797; Conservative 0; Mismatches 3; Indels 4; Gaps 3;

QY 337 GGCAGGTTGACGTGTATGACACCAGCGAGCTCTCTTGTATTTACACATTTGTGCAATCT 396
Db 2 GGCAGGTTGACGTGTATGACACCAGCGAGCTCTCTTGTATTTACACATTTGTGCAATCT 61
QY 397 GAGATTGGGGATTTTGATGAAGCCTTGGACAGAGAGCACTTAGCAAAAATAATACATA 456
Db 62 GAGATTGGGGATTTTGATGAAGCCTTGGACAGAGAGCACTTAGCAAAAATAATACATA 121
QY 457 CCTCAGCAAGCAGCACTAGAGGACAAAATCGTGAATTTCCACATAACCACTTGGACAA 516
Db 122 CCTCAGCAAGCAGCACTAGAGGACAAAATCGTGAATTTCCACATAACCACTTGGACAA 181
QY 517 ACACCAAGCAGAAATCAGATTTTCCAGCTCCTAGAGATTTCCCGCTCGGCTAGAGATGTATGA 576
Db 517 ACACCAAGCAGAAATCAGATTTTCCAGCTCCTAGAGATTTCCCGCTCGGCTAGAGATGTATGA 576

Db	182	ACACCAGCAGAATCAGATTTCCTAGCTCCCTAGAGATTGCCCGTCGGGTAGAGATGTATGGA	241
QY	577	ATCCGGTTGACCCCGCCAGGACAGGAGGACGACGAAGATCAATCTGGCGTTGCGCAAC	636
Db	242	ATCCGGTTGACCCCGCCAGGACAGGAGGACGACGAAGATCAATCTGGCGTTGCGCAAC	301
QY	637	ACGGGAATTCATGTTTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCGCAAGGTG	696
Db	302	ACGGGAATTCATGTTTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCGCAAGGTG	361
QY	697	CGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCGGCGCAGATGCCAATAGT	756
Db	362	CGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCGGCGCAGATGCCAATAGT	421
QY	757	GCCTACAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGG	816
Db	422	GCCTACAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCTGCAAGTCTTCTGG	481
QY	817	AAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAAACGAGCC	876
Db	482	AAATCTGTGTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAAACGAGCC	541
QY	877	AAGCCGCTCTCTTTAGCCGGGGTTCATATTCGCTTCAGTGGTTCGAGTTCAGAGCAG	936
Db	542	AAGCCGCTCTCTTTAGCCGGGGTTCATATTCGCTTCAGTGGTTCGAGTTCAGAGCAG	601
QY	937	GTCTCTGACTATGTT--AAAGAAGGAGGACATAAGAAGTGCAGTTTGAAGGAAGCACA	994
Db	602	GTCTCTGACTATGTTTACACAGAGGAGGACATAAGAAGTGCAGTTTGAAGGAAGCACA	661
QY	995	GCAAGATTCATTTATCCGGAGCGCTTTCACAGCCTTCAGAACTGAAATTCGGAAGTG	1053
Db	662	G-AAGATTCATTTATCCGGAGCGCTTTCACAGCCTTCAGAACTGAAATTCGGAAGTG	720
QY	1054	CTGGACAGTCTCAGCAGAGCACCAGCTTACATTTTGGAGAGGTCCGGAATTCACAGG	1113
Db	721	CTGGACAGTCTCAGCAGAGCACCAGCTTACATTTTGGAGAGGTCCGGAATTCACAGG	780
QY	1114	GGCCAGAGTGCCTGGCGGAGGAAAG	1137
Db	781	GGACAGAGTGCCTGGCGGAGGAAAG	804
RESULT 7			
LOCUS	BE260677	601154188F1 NIH_MGC_19 Homo sapiens	linear EST 26-OCT-2000
DEFINITION		865 bp mRNA clone IMAGE:3510381 5',	
ACCESSION	BE260677	mRNA sequence.	
VERSION	BE260677.1	GI:9132166	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov		
	Plate: LLCM194 row: o column: 22		
	High quality sequence stop: 730.		
FEATURES	Location/Qualifiers		
source	1..865		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		

BASE COUNT	207	a	236	c	255	g	167	t
ORIGIN								
Query Match			32.2%					Score 757.2; DB 10; Length 865;
Best Local Similarity			97.3%					Pred. No. 1.1e-199;
Matches	823;	Conservative	0;	Mismatches	18;	Indels	5;	Gaps 5;
QY	569	TGTATGGAATCCGGTTGCACCCCGCCAGGACAGGAGGACGACGAAGATCAATCTGGCCG	628					
Db	1	TGTATGGAATCCGGTTGCACCCCGCCAGGACAGGACGACGAAGATCAATCTGGCCG	60					
QY	629	TGCCAAACAGGGAATCTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGG	688					
Db	61	TTGCCAACACGGGAATCTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGG	120					
QY	689	CCAAGTGGCGGAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGGCCAGATG	748					
Db	121	CAAGTGGCGGAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGGCCAGATG	180					
QY	749	CCAATAGTCGCTACCAAGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTTCGCAAGT	808					
Db	181	CCAATAGTCGCTACCAAGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTTCGCAAGT	240					
QY	809	CTTCTGGAATAATCTGTTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAC	868					
Db	241	CTTCTGGAATAATCTGTTTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAAC	300					
QY	869	CAAAGCCCAAGCCCGCTCTTTAGCCGGGGTCATCATTTCCGTTTCAGTGGTCCGACTC	928					
Db	301	CAAAGCCCAAGCCCGCTCTTTAGCCGGGGTCATCATTTCCGTTTCAGTGGTCCGACTC	360					
QY	929	AGAAGCAGGTTCTCGACTATGTTTAAAGAGGAGGACATAGAAGTGCAGTTTGAAGGA	988					
Db	361	AGAAGCAGGTTCTCGACTATGTTTAAAGAGGAGGACATAGAAGTGCAGTTTGAAGGA	420					
QY	989	AGCACAGCAAGATTCAATCTATCCGGAGCCTTGTCTTCACAGCCTACAGAACTGAATTCGG	1048					
Db	421	AGCACAGCAAGATTCAATCTATCCGGAGCCTTGTCTTCACAGCCTACAGAACTGAATTCGG	480					
QY	1049	AAGTGTCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGTGCAGAACTCTC	1108					
Db	481	AAGTGTCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGTGCAGAACTCTC	540					
QY	1109	CAGGGGGCCAGAGCTGCCCGCCAGGAAAGAACCCGAAGTTCCTCCGGGGAGCCGGGT	1168					
Db	541	CAGGGGGCCAGAGCTGCCCGCCAGGAAAGAACCCGAAGTTCCTCCGGGGAGCCGGGT	600					
QY	1169	CGACCCGAGCCCTGCGCCGAGGAGAACCCCGCGGGTACAAAGCAGGCGGAGAGCCG	1228					
Db	601	CGACCCGAGCCCTGCGCCGAGGAGAACCCCGCGGGTACAAAGCAGGCGGAGAGCCG	660					
QY	1229	CTTCGGCCCGCCAGGAGGAGGAGGAGTCTTAAAGATAGGACCCAGCAGAGTAAC	1288					
Db	661	CTTCGGCCCGCCAGGAGGAGGAGGAGTCTG-TAAGATAGGA-CCAGCAGAGTTAAC	718					
QY	1289	CTCAGCCCCCGCAGCAACACAGGCTCCCTGACT--GGCAGTCTCTCACCTTTCCGAGCTG	1347					
Db	719	CTCAGCCCCCGCAG-CAAGCAGAGCTCCCTGACTGGGAGCTCTCACCTTTCCGAGCTG	777					
QY	1348	TTTGTGAATTCGAGGGGAGTGGCCCTTGCCAAACGTCGCTTGTCCCAACCTTGAGC	1407					

```
Db 778 TCTTGTGACTCGCAGGGGGAGTTGGCCCTCGCCACAGTAACC-TGTTTCCCAACCTGGGC 836
QY 1408 CCGGAC 1413
    || ||
Db 837 CCGAAC 842

RESULT 8
BG763918          784 bp mRNA linear EST 15-MAY-2001
LOCUS 602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862444 5',
DEFINITION mRNA sequence.
ACCESSION BG763918
VERSION BG763918.1 GI:14074571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 784)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC/DCTD/DTF
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1722 row: 0 column: 21
          High quality sequence stop: 761.
          Location/Qualifiers
            1..784
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4862444"
              /clone_lib="NIH_MGC_49"
              /tissue.type="melanotic melanoma, high MDR (cell line)"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: skin; Vector: pORF7; Site:1: XhoI; Site:2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGCACGAG(G). Size-selected >500bp for average insert size
              1.8kb. Library constructed by Ling Hong in the laboratory
              of Gerald M. Rubin (University of California, Berkeley)
              using ZAP-cDNA synthesis kit (Stratagene) and Superscript
              II RT (Life Technologies). Note: this is a NIH_MGC
              Library. |"
BASE COUNT 203 a 223 c 223 g 135 t
ORIGIN
Query Match 32.1%; Score 755.8; DB 12; Length 784;
Best Local Similarity 99.5%; Pred. No. 2.5e-199;
Matches 779; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1021 GCTTCACAGCCTACAGAACTGAATTCGGAAGTCTGGAGCAGTCTCAGCAGACCAGC 1080
    |||||
Db 2 GCTTCACAGCCTACAGAACTGAATTCGGAAGTCTGGAGCAGTCTCAGCAGACCAGC 61

QY 1081 CTTACATTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGCTGCCGGAGGAAAGGAA 1140
    |||||
Db 62 CTTACATTTGGAGAGGTGCCGAATCTCCAGGGGGCCAGAGCTGCCGGAGGAAAGGAA 121

QY 1141 CCGAAGTTTCCCGCGGGAGCGGGGGTCCGACCCGAGGCCCTCGCGCCGAGGAGAGCCCC 1200
    |||||
Db 122 CCGAAGTTTCCCGCGGGAGCGGGGGTCCGACCCCGAGGCCCTCGCGCCGAGGAGAGCCCC 181

QY 1201 GCGGGTAACAAGCAGCGGAGCGGCGCTCGGCCGCCACGGAGGAAGGAGGAGGTC 1260
    |||||
Db 182 GCGGGTAACAAGCAGCGGAGCGGCGCTCGGCCGCCACGGAGGAAGGAGGAGGTC 241
```

```
QY 1261 GTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCAGCAAGCAGCAGGCTCCCTG 1320
    |||||
Db 242 GTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCGCAGCAAGCAGCAGGCTCCCTG 301

QY 1321 ACTGGCAGTCTCACCCTTTCCGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCG 1380
    |||||
Db 302 ACTGGCAGTCTCACCCTTTCCGAGCTGTCTGTGAATCGCAGGGGGAGTGGCCCTGCG 361

QY 1381 AACGTGACCTTGCTCCCAACCTGAGCCCGCAGCAGCAAGCAGGCTCTCCCTTTGATCAG 1440
    |||||
Db 362 AACGTGACCTTGCTCCCAACCTGAGCCCGCAGCAGCAAGCAGGCTCTCCCTTTGATCAG 421

QY 1441 CGCTGCTGAATGACCAAGGCTGCCCGCGAGCAGCATGAGGATGAGGCGCGAGGAAG 1500
    |||||
Db 422 CGCTGCTGAATGACCAAGGCTGCCCGCGAGCAGCATGAGGATGAGGCGCGAGGAAG 481

QY 1501 AGATTCCTCAACTGATTAAGGCTACTTCATAGCTAAGGAAGTGTCTACCGAGCGAACA 1560
    |||||
Db 482 AGATTCCTCAACTGATTAAGGCTACTTCATAGCTAAGGAAGTGTCTACCGAGCGAACA 541

QY 1561 TATCTGAAGGATCTCGAAGTTATCACTTCGTGCTTTCAGAGCAGCATGAGCAAGAGGAC 1620
    |||||
Db 542 TATCTGAAGGATCTCGAAGTTATCACTTCGTGCTTTCAGAGCAGCATGAGCAAGAGGAC 601

QY 1621 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 1680
    |||||
Db 602 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 661

QY 1681 CATACTAATTTTCTCAAGGAAA-TTGAGCAACGACTTGCCTCTGGGAAGCGCGCTCAA 1739
    |||||
Db 662 CATACTAATTTTCTCAAGGAAA-TTGAGCAACGACTTGCCTCTGGGAAGCGCGCTCAA 721

QY 1740 TGCCCAATCAGAGATTACCAAGAATCGCGATGTCATGC-TGAAGAACAATTCAGGGCA 1798
    |||||
Db 722 TGCCCAATCAGAGATTACCAAGAATCGCGATGTCATGCTTTGACGAAACATTCAGGGCA 781

QY 1799 TGA 1801
    |||
Db 782 TGA 784

RESULT 9
BG762025          1043 bp mRNA linear EST 02-APR-2002
LOCUS AGENCOURT_6859787 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5928511
DEFINITION 5', mRNA sequence.
ACCESSION BG762025
VERSION BG762025.1 GI:19901071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1043)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Rubin Laboratory
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2105 row: c column: 08
          High quality sequence stop: 626.
          Location/Qualifiers
            1..1043
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
FEATURES
          source
```

```
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      266 a   290 c   294 g   193 t
ORIGIN
Query Match      31.7%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 3.5e-196;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1591 TGGTTTCAGAGCAGCTGAGCAAGAGGAGCGCATCCGGAAGCAGCTGAAAAGTCTCATA 1650
DB 1 TGGTTTCAGAGCAGCTGAGCAAGAGGAGCGCATCCGGAAGCAGCTGAAAAGTCTCATA 60
QY 1651 TTCCCGAATTTTGAACCTTTGACAAAATTTTCATACATAATTTTCTCAAGGAAATTGAGCAA 1710
DB 61 TTCCCGAATTTTGAACCTTTGACAAAATTTTCATACATAATTTTCTCAAGGAAATTGAGCAA 120
QY 1711 CGACTTGCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCG 1770
DB 121 CGACTTGCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCG 180
QY 1771 GATGTCATGCTGAAGAAACATTCAGGCGCATGAAGCAGCTCGCGGCTCACCTGTGGGAAGCAC 1830
DB 181 GATGTCATGCTGAAGAAACATTCAGGCGCATGAAGCAGCTCGCGGCTCACCTGTGGGAAGCAC 240
QY 1831 AGCGAGGCTTTGGAGCCCTGGAGAAATGAATCAAGAGCTCCCGGCGGTGAGGAACCTTC 1890
DB 241 AGCGAGGCTTTGGAGCCCTGGAGAAATGAATCAAGAGCTCCCGGCGGTGAGGAACCTTC 300
QY 1891 TGCAGAGACTTTCAGCTGAGAGGTGTTTACCTACCCCTCAACACCTTCCTCTCGGCG 1950
DB 301 TGCAGAGACTTTCAGCTGAGAGGTGTTTACCTACCCCTCAACACCTTCCTCTCGGCG 360
QY 1951 CCAGTCCACCGGCTCATGCACTACAAGCAGGTCTCGAGCGGCTGTGCAAAACACACCGG 2010
DB 361 CCAGTCCACCGGCTCATGCACTACAAGCAGGTCTCGAGCGGCTGTGCAAAACACACCGG 420
QY 2011 CCGAGCCACCGGCTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTG 2070
DB 421 CCGAGCCACCGGCTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTG 480
QY 2071 GCACAGCTCCACGGTACGATGATCAGATGGAGAAATTCAGAGCTGCACGAACCTCAG 2130
DB 481 GCACAGCTCCACGGTACGATGATCAGATGGAGAAATTCAGAGCTGCACGAACCTCAG 540
QY 2131 AAAGATTTCATTGGCAATTCGACAAATCTTGTTGGTTCCGGGAAGGAGTTTCATCCGCTGGGC 2190
DB 541 AAAGATTTCATTGGCAATTCGACAAATCTTGTTGGTTCCGGGAAGGAGTTTCATCCGCTGGGC 600
QY 2191 AGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCACAGAC 2250
DB 601 AGCCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCCTGTTCACAGAC 660
QY 2251 GTCCTGCTATACAGCGCGGGGCTGACGGGCTCCAATCAGTTTAAAGTCCACGGGCGAG 2310
DB 661 GTCCTGCTATACAGCGCGGGGCTGACGGGCTCCAATCAGTTTAAAGTCCACGGGCGAG 720
QY 2311 CTCGCCCTCTATGGCATGACGATTGAGG 2338
DB 721 CTCGCCCTCTATGGCATGACGATTGAGG 748
```

```
RESULT 10
LOCUS      BG475554
DEFINITION BG475554 901 bp mRNA linear EST 21-MAR-2001
            602491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
            mRNA sequence.
ACCESSION  BG475554
VERSION    BG475554.1 GI:13407833
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI381 row: h column: 11
            High quality sequence stop: 818.
            Location/Qualifiers
                1..901
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4620346"
                /clone_lib="NIH_MGC_20"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 232 a 243 c 253 g 173 t

ORIGIN

Query Match 30.9%; Score 726.2; DB 12; Length 901;

Best Local Similarity 91.3%; Pred. No. 5e-191;

Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;

QY 1430 CCTTGATCAGCCGCTGCTGAATGACCGGCTGCCCGCGGACGAGATGAGGATGAGG 1489

DB 2 CCTTGATCAGCCGCTGCTGAATGACCGGCTGCCCGCGGACGAGATGAGGATGAGG 61

QY 1490 GCCGAGGAGAGATTCGCCAAGCTGATAAGCGCTACTTACCTAGCTAAGGAAGTGTCTACCA 1549

DB 62 GCCGAGGAGAGATTCGCCAAGCTGATAAGCGCTACTTACCTAGCTAAGGAAGTGTCTACCA 121

QY 1550 CCGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTTGTTTCAGACGACAGTGA 1609

DB 122 CCGAGCGAATATCTGAAGGATCTCGAAGTTATCACTTCGTTGTTTCAGACGACAGTGA 181

QY 1610 GCAAGAGAGACGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTT 1669

DB 182 GCAAGAGAGACGCCATGCCGGAAGCACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTT 241

QY 1670 TGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTGGGAAG 1729

DB 242 TGCACAAATTTTCATACATAATTTTCTCAAGGAAATTTGAGCAACGACTTGCCTGTGGGAAG 301

QY 1730 GCCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTCATGCTGCTGAAGAACA 1789

DB 302 GCCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTCATGCTGCTGAAGAACA 361


```
QY 1790 TTCAGGCGATGAAGCACTCGCGGCTCACCTGTGGAAGCACACGAGGCGCTTGAGGCC 1849
|||||
Db 362 TTCAGGCGATGAAGCACTCGCGGCTCACCTGTGGAAGCACACGAGGCGCTTGAGGCC 421
QY 1850 TGCAGATGAATCAAGAGCTCCGCGGCTGGAGACTTCTGCAGAGACTTTGAGCTGC 1909
|||||
Db 422 TGAGATGAATCAAGAGCTCCGCGGCTGGAGACTTCTGCAGAGACTTTGAGCTGC 481
QY 1910 AGAAGTGTGTTACCTACCGCTCAACACCTTCCTCTCGCGGCTGCACGCTCATGC 1969
|||||
Db 482 AGAAGTGTGTTACCTACCGCTCAACACCTTCCTCTCGCGGCTGCACGCTCATGC 541
QY 1970 ACTAAGCAGGTCCTGGAGCGCTGTGCAACACCAACCGCGGAGCGCGGACTTCA 2029
Db 542 ACTAAGCAGGTCCTGGAGCGCTGTGCAACACCAACCGCGGAGCGCGGACTTCA 601
QY 2030 -GGACTCGCGAGCGCTTGGCAGAGATCACGAGATGTTGTCACAGCTCCACGGTAG 2088
Db 602 GGGACTGGCGAGCGCTTGGCAGAGATCACGAGATGTTGTCACAGCTCCACGGTAG 661
QY 2089 ATGATCAAGATGAGAAATTCACAGAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATT 2148
|||||
Db 662 ATGATCAAGATGAGAAATTCACAGAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATT 721
QY 2149 GACAACTTGTGTTCGGGAA----GGGAGTTTCATCCGCTCTGGGAGCGCTCAGCAA--- 2201
|||||
Db 722 GACAACTTGTGTTCGGGAA----GGGAGTTTCATCCGCTCTGGGAGCGCTCAGCAAAGC 781
QY 2202 -GCTCTCGGGAGGGGCTCCACAGCGCATGTTCTTCTGTTCAAGGAGTCTCT-GCTA 2259
|||||
Db 782 TCCTCGGGGAGGGGCTCCACAGCGCGCATGTTCTTCTGTTCAAGGAGTCTCTGGCTT 841
QY 2260 TACACGAGCGGGGCTGACGGGCTCCATCAGTTTAAAGTCCACGGGAGCTCC 2314
|||||
Db 842 TACACGAGCGGGGCTGACAGGGCTCCATCCAGTTTAAAGCCCAAGGCGAGTCCC 896

RESULT 11
BE745887 898 bp mRNA linear EST 15-SEP-2000
LOCUS 601573513F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3834699 5',
DEFINITION mRNA sequence.
ACCESSION BE745887
VERSION BE745887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 898)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LNCM515 row: i column: 04
High quality sequence stop: 723.
Location/Qualifiers
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3834699"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:

```

```
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).*
BASE COUNT 236 a 229 c 269 g 164 t
ORIGIN
Query Match 30.0%; Score 705.4; DB 12; Length 898;
Best Local Similarity 96.8%; Pred. No. 3.2e-185;
Matches 762; Conservative 0; Mismatches 21; Indels 4; Gaps 4;
QY 511 GGACAAACACACGAGAAATTCACAGTTCCTCAGATTCGCCCTCGGCTAGAGATG 570
|||||
Db 1 GGACAAACACACGAGAAATTCACAGTTCCTCAGATTCGCCCTCGGCTAGAGATG 60
QY 571 TATGGAATCCGTTGACCCGCAAGGACAGGAGGACGAAGATCAATCTGCGCGTT 630
|||||
Db 61 TATGGAATCCGTTGACCCGCAAGGACAGGAGGACGAAGATCAATCTGCGCGTT 120
QY 631 GCCAACACGGAATTCAGTGTTCAGGTTTCTACTAAGATCAATGCTTCAACTGGGCC 690
|||||
Db 121 GCCAACACGGAATTCAGTGTTCAGGTTTCTACTAAGATCAATGCTTCAACTGGGCC 180
QY 691 AAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGCCAGATGCC 750
|||||
Db 181 AAGGTGCGGAAGCTGAGCTTCAAGAGGAGCGCTTCTCATCAAGCTCCGCCAGATGCC 240
QY 751 AATAGTCCGTACCAGGATACCTTTGGAATTCCTGATGCCAGTGGGATTTCTCAAGTCC 810
|||||
Db 241 AATAGTCCGTACCAGGATACCTTTGGAATTCCTGATGCCAGTGGGATTTCTCAAGTCC 300
QY 811 TTCTGGAATCTGTTGAAACATCATGCCCTCTTTAGACTTTTGAAGAGCCCAACCA 870
|||||
Db 301 TTCTGGAATCTGTTGAAACATCATGCCCTCTTTAGACTTTTGAAGAGCCCAACCA 359
QY 871 AAGCCCAAGCGCTCTCTTTAGCGGGGTCTATCTTTTCGTTTCAGTGTTCGAGCTCAG 930
|||||
Db 360 AAGCCCAAGCGCTCTCTTTAGCGGGGTCTATCTTTTCGTTTCAGTGTTCGAGCTCAG 419
QY 931 AAGCAGGTTCTCGACTATGTTAA-AGAAGGAGGACATTAAGAAGGTGCAGTTTGAAGAGGAA 989
|||||
Db 420 AAGCAGGTTCTCGACTATGTTAA-AGAAGGAGGACATTAAGAAGGTGCAGTTTGAAGAGGAA 479
QY 990 GCACAGCAAGATTTCATTCTATCCGGAGCCTTGTCTCACACCTTACAGAACTGAATTCGGA 1049
|||||
Db 480 GCACAGCAAGATTTCATTCTATCCGGAGCCTTGTCTCACAGCCTTACAGAACTGAATTCGGA 539
QY 1050 AGTGCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTTGGAGAGGTCGCCAATCTCC 1109
|||||
Db 540 AGTGCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTTGGAGAGGTCGCCAATCTCC 599
QY 1110 AGGGGCCAGAGTGTCCGCGAGGAAAGGAAACCGAGTTTCCGCCGGGAGCGCGGGTC 1169
|||||
Db 600 A-GGGGCCAGAGTGTCCGCGAGGAAAGGAAACCGAGTTTCCGCCGGGAGCGCGGGTC 658
QY 1170 GCACCCGAGCCCTGCGCGGAGGAGAGAGCCCGCGGGTAAACAAGCAGCGGACGAGCCGC 1229
|||||
Db 659 GCACCCGAGCCCTGCGCGGAGGAGAGAGCCCGCGGGTAAACAAGCAGCGGACGAGCCGC- 717
QY 1230 CTCGGCGCCACGAGGAGGAGGAGGAGGTCGTTAAGGATAGACCCAGCAGAGTAAACC 1289
|||||
Db 718 CTCGGCGCCACGAGGAGGAGGAGGAGGCGCCGTTAAGTAGGAGCCCAAGAGAGTAAAC 777
QY 1290 TCAGCCC 1296
|||||
Db 778 TTAGCCC 784
RESULT 12
BG829192
```

LOCUS BG829192 835 bp mRNA linear EST 22-MAY-2001
DEFINITION 602753246F1 NTH_MGC_17 Homo sapiens cDNA clone IMAGE:4906114 5',
mRNA sequence.
ACCESSION BG829192
VERSION BG829192.1 GI:14176766
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 835)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1808 row: k column: 11
High quality sequence stop: 835.
Location/Qualifiers
1. 835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906114"
/clone_lib="NTH_MGC_17"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 213 a 231 c 243 g 148 t
ORIGIN
Query Match 30.0%; Score 704.6; DB 12; Length 835;
Best Local Similarity 95.8%; Pred No. 5.1e-185;
Matches 800; Conservative 0; Mismatches 24; Indels 11; Gaps 7;
QY 826 GTTGAACATCATGCGCTCTTTAGACTTTTGAAGAGCCCCAACCAAGCCCCAGCCCGTC 885
Db 2 GTTGAACATCATGCGCTCTTTA-ACITTTTGAAGAGCCCCAACCAAGCCCCAGCCCGTC 60
QY 886 CTTCTTAGCCGGGGTCATCATTTCCGTTTCAGTGTGCGACTCAGAAGCAGGTTCTCGAC 945
Db 61 CTTCTTAGCCGGGGTCATCATTTCCGTTTCAGTGTGCGACTCAGAAGCAGGTTCTCGAC 120
QY 946 TATGTTAAAGAGGAGGACATAGAAGGTGCAGTTTGAAGGAAGCACACAGCAAGATTTCAT 1005
Db 121 TATGTTAAAGAGGAGGACATAGAAGGTGCAGTTTGAAGGAAGCACACAGCAAGATTTCAT 180
QY 1006 TCTATCCGGAGCGCTTCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCT 1065
Db 181 TCTATCCGGAGCGCTTCTTCACAGCCTACAGAACTGAATTCGGAAGTGTGGAGCAGTCT 240
QY 1066 CAGCAGACACAGCGCTTACATTGAGAAAGTGCCGAATCTCCAGGGGGCCAGAGCTGC 1125
Db 241 CAGCAGACACAGCGCTTACATTGAGAAAGTGCCGAATCTCCAGGGGGCCAGAGCTGC 300
QY 1126 CGCGAGGAAAGCAACCGAGGTTTCCGCCGGGAGCCGGGTGCGACCCGAGCCCTCGG 1185
Db 301 CGCGAGGAAAGCAACCGAGGTTTCCGCCGGGAGCCGGGTGCGACCCGAGCCCTCGG 360
QY 1186 CCGAGGAGAGCCCGCGGGTAAACAGCAGCGGAGCGCGCTCGCGGCCCGCCAGGAG 1245

Db 361 CCGAGGAGAAGCCCGCGGGTAAACAAGCAGCGGACGGAGCGCGCTCGCGGCCCGCAGGAG 420
QY 1246 GAAGAGGAGGAGTGGTT-AAGGATAGGACCCAGCAGAGAGTAACTTCAGCCCCCGCAGCC 1304
Db 421 GAAGAGGAGGAGTGGTTCAAGGATAGGACCCAGCAGAGTAACTTCAGCCCCCGCAGCC 480
QY 1305 AAGCAGAGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAATCGCAGGG 1364
Db 481 AAGCAGAGCTCCCTGACTGGCAGTCCCTCACCTTTCCGAGCTGTCTGTGAATCGCAGGG 540
QY 1365 GGGAGTGGCCCCCTGCCAACGTGACCTTGTCTCCCAACTGAGCCCCCGCAGCAACGAGCC 1424
Db 541 GGGAGTGGCCCCCTGCCAACGTGACCTTGTCTCCAGACTGAGCCCCGAAACCAAGCAGCC 600
QY 1425 CTTCTCCCTTGATCAGCCCGCTGC-TGAATGACCAAGGCTGCCCGGAGCGGAGGATGAG 1483
Db 601 CTTCTCCCTTGATCAGCCCGCTGTTGAATGACCAAGGCTGCCCGGAGCGGAGGATGAG 660
QY 1484 ATGAGGGCCGGAGGAGAGATTCCCAACTGATAAGCGTACTTCA-TAGCTAAGGAAGTG 1542
Db 661 ATGAGGGCCGGAGGAGAGATTCCCAACTGATAAGCGTACTTCA-TAGCTAAGGAAGTG 720
QY 1543 TCTACCACCGAGCGAAC---ATATCTGAAAGGATCTCGAAG-TTATCACTTCGTGTTTCA 1598
Db 721 TCTACCACCGAGCGAACCATATTCTGAAGGATCTCGAAGTTTATCACTTCGCGGTTCAG 780
QY 1599 GAGCAGACTGACCAAGAGGAGCGCATGCCGGAAGCAC---TGAAAAGTCTCATATA 1650
Db 781 AAGCAGACTGAAAGCAAGAGGAGCGCATGCCGGAAGCACCTTGAAAAGTCTCATATA 835
RESULT 13
BE910036
LOCUS 601498085F1 NTH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
mRNA sequence.
DEFINITION BE910036
ACCESSION BE910036
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 902)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9498 row: j column: 22
High quality sequence stop: 750.
Location/Qualifiers
1. 902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NTH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN

Query Match		29.8%;	Score 700.2;	DB 12;	Length 902;					
Best Local Similarity		95.8%;	Pred. No. 9e-184;							
Matches 752;		Conservative 0;	Mismatches 28;	Indels 5;	Gaps 3;					
QY	1041	GAATTCGGAGTCTCGGACAGTCTCAGCAGACACACAGCCTTACATTTGGAGAAGTGC	1100							
Db	1	GAATTCGGAGTCTCGGACAGTCTCAGCAGACACACAGCCTTACATTTGGAGAAGTGC	60							
QY	1101	CGAATCTCCAGGGGCCAGAGCTGCCGGCGAGGAAAGAACCGAAGGTTTCCGCCGGGA	1160							
Db	61	CGAATCTCCA-GGGCCAGAGCTGCCGGCGAGGAAGAACCGAAGTTCGCCCGGGGA	119							
QY	1161	GCCGGGTGCAACCCAGGCGCTGCGCCGAGGAGAAGCCCGCGGTAAACAAGCAGCGGA	1220							
Db	120	GCCGGGTGCAACCCAGGCGCTGCGCCGAGGAGAAGCCCGCGGTAAACAAGCAGCGGA	179							
QY	1221	CGAGCGCGCTCGCGCCACGAGGAGGAGGAGTTCGTTAAGATAGGACCCAGCA	1280							
Db	180	CGAGCGCGCTCGCGCCACGAGGAGGAGGAGTTCGTTAAGATAGGACCCAGCA	239							
QY	1281	GAGTAAACCTCAGCCCGCAGCAAGCAGGCTCCCTGACTGGCAGTCTCACCTTTC	1340							
Db	240	GAGTAAACCTCAGCCCGCAGCAAGCAGGCTCCCTGACTGGCAGTCTCACCTTTC	299							
QY	1341	CGAGCTGTCTGTGAACTCGACAGGGGGAGTGGCCCTGCCAACGTCGACCTGTCTCCAA	1400							
Db	300	CGAGCTGTCTGTGAACTCGACAGGGGGAGTGGCCCTGCCAACGTCGACCTGTCTCCAA	359							
QY	1401	CTTGAGCCCGGACACCAAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAAGC	1460							
Db	360	CTTGAGCCCGGACACCAAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAAGC	419							
QY	1461	CTGCCCGCGGACGAGGATGAGGCGCGGAGGAGAGATTCCCAACTGATAAAGC	1520							
Db	420	CTGCCCGCGGACGAGGATGAGGCGCGGAGGAGAGATTCCCAACTGATAAAGC	479							
QY	1521	GTACTTATAGCTAAGGAAGTGTCTACCGCGGAGCAATATCTCTGAAGGATCTCGAAGT	1580							
Db	480	GTACTTATAGCTAAGGAAGTGTCTACCGCGGAGCAATATCTCTGAAGGATCTCGAAGT	539							
QY	1581	TATCACTTCTGGTTTCAGACGACAGTGAAGAAAGAGGAGCGCATGCCGG-AAGCACTGA	1639							
Db	540	TATCACTTCTGGTTTCAGACGACAGTGAAGAAAGAGGAGCGCATGCCGGAAAGCACTGA	599							
QY	1640	AAA---GTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTTCTACTAATTTCTCA	1696							
Db	600	AAAAGTCTCATATTTCCCGAATTTTGAACCTTTGCACAAATTTCTACTAATTTCTCA	659							
QY	1697	AGAAATGAGCAACGACTTGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATT	1756							
Db	660	AGAAATGAGCAACGACTTGCCCTGTGGGAAGGCGCTCAATGCCCAATCAGAGATT	719							
QY	1757	ACCAAGAATCGCGATGTCATGCTGAAGAACAATTCAGGCGCATGAAGACGCTGCGCGGCTC	1816							
Db	720	ACCAAGAATCGCGATGTCATGCTGAAGAACAATTCAGGCGCATGAAGACGCTGCGCGGCTC	779							
QY	1817	ACCTG 1821								
Db	780	CTGTG 784								
RESULT 14										
AUI32546										
LOCUS										
DEFINITION										
AUI32546 NT2RP4 Homo sapiens cDNA clone NT2RP4000050 5', mRNA										
ACCESSION										
VERSION										
AUI32546.1 GI:10992900										
KEYWORDS										
EST.										
SOURCE										
human.										
ORGANISM										
Homo sapiens										
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										

REFERENCE		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS		1 (bases 1 to 721) Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and Isogai,T.
TITLE		HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano ,S., Masuho,Y., Isogai,T.)
JOURNAL		Unpublished (2000)
COMMENT		Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES		Location/Qualifiers
SOURCE		1..721 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RP4000050" /cell_lib="NT2RP4" /cell_type="teratocarcinoma" /cell_line="NT2" /note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction" BASE COUNT 183 a 180 c 214 g 141 t 3 others ORIGIN
Query Match		29.2%; Score 687.6; DB 9; Length 721;
Best Local Similarity		98.8%; Pred. No. 2.6e-180;
Matches 712;		Conservative 0; Mismatches 7; Indels 2; Gaps 2;

```
QY 1099 GCGAATCTCCAGGGGCGAGAGTCCCGGCGAGGAAGAAACCGAAGGTTTCCCGCGGG 1158
|||||
Db 541 GCGAATCTCCAGGGGCGAGAGTCCCGGCGAGGAAGAAACCGAAGGTTTCCCGCGGG 600
|||||
QY 1159 GAG-CCGGGGTGCACCCGAGCCCTGCGCGGAGGAGAACCCCGGGGTAAACAAGCAGGC 1217
|||
Db 601 GAGCCGGGGTGCACCCGAGCCCTGCGCGGAGGAGAACCCCGGGGTAAACAAGCAGGC 660
|||||
QY 1218 GGACGAGCGCCCTCGGCCCGCCACGAGGAAGAGGAGGAGTGC-TTAAGGATAGGACCC 1276
|||||
Db 661 GGACGAGCGCCCTGNGCGCCCGCCACGAGGAAGAAGANGANGTCGTTAAGGATAGGACCC 720
|||||
QY 1277 A 1277
|
Db 721 A 721

RESULT 15
LOCUS BQ674703
DEFINITION AGENCOURT_818B989 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255217
5', mRNA sequence.
ACCESSION BQ674703
VERSION BQ674703.1 GI:21785537
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 988)
AUTHORS NIH-MGC http://mgc.mci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2405 row: p column: 02
High quality sequence stop: 685.
FEATURES
Location/Qualifiers
1..988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6255217"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
Synthesis Kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 216 a 306 c 291 g 161 t 14 others
ORIGIN
Query Match 29.0%; Score 682; DB 14; Length 988;
Best Local Similarity 99.6%; Pred. No. 1.2e-178;
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1657 AATTGTGAACCTTTGCACAAATTTATATACTAATTTCTCAAGAAATTCAGCAACGACTT 1716
|||||
Db 1 AATTGTGAACCTTTGCACAAATTTATATACTAATTTCTCAAGAAATTCAGCAACGACTT 60
|||||
QY 1717 GCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 1776
|||||
```

```
Db 61 GGCCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAATCGCGGATGTC 120
QY 1777 ATGCTGAAGAACATTCAGGGCATGAGCACCTGGCGGCTCACCTGTGGAAGCACACCGAG 1836
|||||
Db 121 ATGCTGAAGAACATTCAGGGCATGAGCACCTGGCGGCTCACCTGTGGAAGCACACCGAG 180
|||||
QY 1837 GCCTTGGAGCGCCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGA 1896
|||||
Db 181 GCCTTGGAGCGCCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGA 240
QY 1897 GACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCGGGGCACTG 1956
|||||
Db 241 GACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCGGGGCACTG 300
QY 1957 CACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCGCGGAGC 2016
|||||
Db 301 CACCGGCTCATGCACACTACAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCGCGGAGC 360
QY 2017 CACGCGGACTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACAG 2076
|||||
Db 361 CACGCGGACTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACAG 420
QY 2077 CTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCAGAACTCAAGAAAGAT 2136
|||||
Db 421 CTCCACGGTACGATGATCAAGATGGAGAATTTCCAGAAGCTGCAGAACTCAAGAAAGAT 480
QY 2137 TTGATTGGCATTTGACAATCTTGTGGTTCCGGGAAGGGAGTTTCATCCGTTGGGCGAGCTC 2196
|||||
Db 481 TTGATTGGCATTTGACAATCTTGTGGTTCCGGGAAGGGAGTTTCATCCGTTGGGCGAGCTC 540
QY 2197 AGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 2256
|||||
Db 541 AGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTG 600
QY 2257 CTATACAGAG-CCGGGGGCTGACGGCCCTCCAATCAGTTTAAAGTCCACGGGAGCTCC 2315
|||||
Db 601 CTATACAGAGNCCGGGGGCTGACGGCCCTCCAATCAGTTTAAAGTCCACGGGAGCTCC 660
QY 2316 GCTCTATGCATGAGCATGTGAGGAGCGCAAGACGA 2351
|||||
Db 661 GCTCTATGCATGAGCATGTGAGGAGCGCAAGACNA 696

Search completed: December 3, 2002, 21:25:01
Job time : 2576.22 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 3, 2002, 15:35:00 ; Search time 496.761 Seconds
(without alignments)
16520.996 Million cell updates/sec
Title: US-09-555-342B-1_COPY_151_432
Perfect score: 282
Sequence: 1 ccttcagaaactgtgtc.....ctgaccacacaaactccaa 282

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	282	100.0	3442	9	AB008430	AB008430 Homo sapi
2	107	37.9	96183	9	AL161896	AL161896 Homo sapi
3	107	37.9	152053	2	AC022669	AC022669 Homo sapi
4	97.8	34.7	2605	9	BC021301	BC021301 Homo sapi
5	97.8	34.7	3997	9	AB018336	AB018336 Homo sapi
6	92.2	32.7	3719	10	BC009153	BC009153 Mus muscu
7	75	26.6	149417	2	AC118779	AC118779 Rattus no
8	73.4	26.0	221502	2	AC122885	AC122885 Mus muscu
9	71.6	25.4	2882	10	AF106702	AF106702 Mus muscu
10	70.8	25.1	170916	9	AL136300	AL136300 Human DNA
11	70.4	25.0	3451	10	AB032366	AB032366 Mus muscu
12	68.4	24.3	3619	9	BC010674	BC010674 Homo sapi
13	68.4	24.3	3643	9	HUMPTYPH	M68941 Human prote
14	65.4	23.2	152053	2	AC022669	AC022669 Homo sapi
15	65.2	23.1	2308	9	AB032179	AB032179 Homo sapi
16	65.2	23.1	3313	6	AX050013	AX050013 Sequence
17	65.2	23.1	3770	9	AF153418	AF153418 Homo sapi
18	65.2	23.1	5582	9	AF153416	AF153416 Homo sapi
19	62.4	22.1	62537	2	AC101100	AC101100 Mus muscu
20	61.4	21.8	14917	6	AX375624	AX375624 Sequence
21	60.6	21.5	3106	6	AX050012	AX050012 Sequence
22	60	21.3	5097	5	AY124488	AY124488 Danio rer
23	59.8	21.2	224310	2	AC094777	AC094777 Rattus no
24	59.4	21.1	2595	9	AF156225	AF156225 Homo sapi
25	58.2	20.6	2964	10	AF044312	AF044312 Mus muscu
26	57.8	20.5	621	6	AX430372	AX430372 Sequence
27	57.8	20.5	2835	9	BC007796	BC007796 Homo sapi
28	57.8	20.5	2867	9	HUMELIA	M14993 Human struc
29	57.8	20.5	3043	9	HUMEMP41	M61733 Homo sapien
30	57	20.2	2717	10	AF177146	AF177146 Mus muscu
31	57	20.2	3661	10	AB032828	AB032828 Rattus no
32	57	20.2	4051	10	AF152247	AF152247 Mus muscu
33	57	20.2	4543	10	AB032827	AB032827 Rattus no
34	56.8	20.1	2625	9	AY049789	AY049789 Homo sapi
35	56.8	20.1	3534	9	BC013885	BC013885 Homo sapi
36	56.8	20.1	6263	9	AB002336	AB002336 Human mRN
37	56.6	20.1	2970	9	AK000388	AK000388 Homo sapi
38	55.2	19.6	5855	10	AB019257	AB019257 Rattus no
39	55.2	19.6	6102	10	AB019256	AB019256 Rattus no
40	55	19.5	1784	9	AK094281	AK094281 Homo sapi
41	54.6	19.4	3984	9	HUMCAP	M64572 Human prote
42	54.6	19.4	5942	10	MUS41SP	L00919 Mus musculu
43	54	19.1	3166	9	AK023019	AK023019 Homo sapi
44	53.4	18.9	2758	5	XELCSK	M20621 X.laevlis cy
45	52.8	18.7	3872	10	BC034751	BC034751 Mus muscu

ALIGNMENTS

RESULT 1
AB008430
LOCUS AB008430 Homo sapiens mRNA for CDEP, complete cds.
DEFINITION Homo sapiens mRNA for CDEP, complete cds.
ACCESSION AB008430
VERSION AB008430.1 GI:2766164
KEYWORDS CDEP.
SOURCE Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and Kato,Y.
TITLE Molecular cloning and characterization of CDEP, a novel human

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors

Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)

98086358

2 (bases 1 to 3442)

Koyano,Y., Kawamoto,T. and Kato,Y.

Direct Submission

Submitted (22-OCT-1997) Takeshi Kawamoto, Hiroshima University

School of Dentistry, Department of Biochemistry; 1-2-3 Kasumi

Minami-ku, Hiroshima, Hiroshima 734, Japan

(E-mail:tkawano@ipc.hiroshima-u.ac.jp. Tel:082-257-5688,

Fax:082-257-5629)

FEATURES

Location/Qualifiers

1..3442

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="chondrocyte"

/tissue_type="cartilage"

/dev_stage="embryo"

49..3186

/function="Rho Guanine Nucleotide Exchange Factor"

/note="Band 4.1 superfamily"

/codon_start=1

/product="CDEP"

/protein_id="BAA24267.1"

/db_xref="GI:2766165"

/translation="MGEIQRTPGSRIGAPENSIGSTLERGOKPPPTPSKLVSIKI
QMDDQFAFVQAPKQVLDVAVCNHLNVGDFGLFPPDHRKRTTWLIDLKPIV
KOIRDPKVVVYKFFPPDHTQLEELRYFALQVKQDLAQGLRCLNTDSALLI
SHVQSEIGDDEALDRHLAKNKYIPQDDALEDKIVEFHNNHIGQTPAESDFQLLEI
ARLLMEYGLRLHPKDRGRTKINLAVANTGILVFOGFTKINAFNAKVKLSFKRKR
LTKLPDANSAYQDVFLEFLMASRDFCKSEFWKVEHFAFERFEPKPKPKPVLFSRG
SSEFSGTKQVLDYVKEGKKVKQFEPKHSKIHISIRLSAQPELSEVLEQSQQS
TSLTGEAESPGGSGRGKPKVSAGEPGSHPSAPRPSAGNKGADGAASAPTEE
EEVYKDTQSKQPPQPPSTGSLTSGHLSVNSQGGVAPANVTLSNLSPTDKQ
ASPLPLNDQACPTDEDEGRKRFPTDKAYIAKEVSTTERTYTLKDLVITSWF
QSTVSKEDAMPALAKLIFPNPEPLHKFTNFKIEQRLALWEGRSNAQIRDYORIG
DVMKLNQGMKHLAAHLKHSLEALENGIKSSRLENFCRDFLOKVCYLPNTFL
LRLHLHYKQVLRCKHPPHSHADPRDCRAALAEITEMVAOLHGMIMKRNPKL
HELKDLGIDNLVPGREFRLGSLKSGKGLQORMPFLFNVLTLTSGRLTASNQ
FKVGLPLGIMTIESEDEWGPVHCLTLRQSGSLIIVAASRSMERKWEVIEDQNAID
LAEKSSPAPELSPDNKSPDEATAADQESDLSASRTSLERQAPHRGNTMVHP
CWHRTSVSWDFSLAVNQSLNLRKFNKSNQWKLWVVTNFCFLFFYKSHQDNHP
LASPLGLSLTIPSESENIQKDYVFKLHFKSHVYFRAESEYTPFERMWEVIRSATSS
ASRPHVLSKESLVY"

3442

/note="50 a nucleotides"

BASE COUNT 864 a 952 c 927 g 699 t

polya_site

BASE COUNT 864 a 952 c 927 g 699 t

Query Match 100.0%; Score 282; DB 9; Length 3442;

Best Local Similarity 100.0%; Pred. No. 1.4e-77;

Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGAAAACCTGTGCCATCAAAATCCAGATCCTGGATGACACACCGAGGCGATT 60

Db 151 CCTTCAGAAAACCTGTGCCATCAAAATCCAGATCCTGGATGACACACCGAGGCGATT 210

QY 61 GAAGTTCCCAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC 120

Db 211 GAAGTTCCCAAGAGCTCCTGGGAAGGTGCTGCTGGATGCAGTTTGCACCAACCTCAAC 270

QY 121 CTCGTGGAAGGTGACTATTTCGCCCTCGAGTTTCCTGATGACAAAAAGATCAGGTGCG 180

Db 271 CTCGTGGAAGGTGACTATTTCGCCCTCGAGTTTCCTGATGACAAAAAGATCAGGTGCG 330

QY 181 CTGATCTCCTAAACCCATTGTGAACAGATTAGAGGCCAACGACCGTTGTTGTTAAG 240

Db 331 CTGATCTCCTAAACCCATTGTGAACAGATTAGAGGCCAACGACCGTTGTTGTTAAG 390

QY 241 TTGTGGTGAATTCCTCCGCTGACACACCAACTCCAA 282

Db 391 TTGTGGTGAATTCCTCCGCTGACACACCAACTCCAA 432

RESULT 2

AL161896

LOCUS

DEFINITION

complete sequence.

ACCESSION

AL161896

VERSION

AL161896.16

GI:12330752

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 96183)

Direct Submission

Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jan 22, 2001 this sequence version replaced gi:12329254.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

Rpl1-261P24 is from the library RPCI-11.1 constructed by the group

of Peter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone

Rpl1-261P24. It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone Rpl1-261P24 is at 96183 in this

sequence. The true left end of clone Rpl1-31K22 is at 59428 in this

sequence. The true right end of clone Rpl1-573N10 is at 100 in this

sequence.

FEATURES

Location/Qualifiers

1..96183

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="13"

/clone="Rpl1-261P24"

/clone_lib="RPCI-11.1"

1301..1335

/note="SN5 repeat: matches 34..68 of consensus"

1432..1506

/note="MER20 repeat: matches 1..75 of consensus"

3518..3854

/note="L1MB8 repeat: matches 5837..6173 of consensus"

5528..5599

/note="12 copies 6 mer tatatta 69% conserved"

5533..5604

/note="36 copies 2 mer at 69% conserved"

5766..6068

/note="AluSp repeat: matches 7..303 of consensus"

6949..7264

repeat_region /note="AluJo repeat: matches 1. .307 of consensus"
7380. .7501
repeat_region /note="MIR repeat: matches 65. .191 of consensus"
8192. .8407
repeat_region /note="L2 repeat: matches 2064. .2289 of consensus"
8475. .8506
repeat_region /note="L6 copies 2 mer tt 100% conserved"
8929. .9103
repeat_region /note="MER33 repeat: matches 4. .171 of consensus"
9104. .9421
repeat_region /note="AluJb repeat: matches 1. .306 of consensus"
9422. .9450
repeat_region /note="MER33 repeat: matches 171. .216 of consensus"
9491. .9595
repeat_region /note="MER33 repeat: matches 142. .225 of consensus"
9596. .9909
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
9910. .9997
repeat_region /note="MER33 repeat: matches 225. .319 of consensus"
11665. .11763
repeat_region /note="MIR repeat: matches 47. .149 of consensus"
11991. .12022
repeat_region /note="8 copies 4 mer tgtg 90% conserved"
12274. .12558
repeat_region /note="AluX repeat: matches 20. .311 of consensus"
12577. .12879
repeat_region /note="AluY repeat: matches 1. .304 of consensus"
13406. .13549
repeat_region /note="MLTIC repeat: matches 1. .146 of consensus"
13554. .13595
repeat_region /note="L6 copies 2 mer ca 96% conserved"
13682. .13715
repeat_region /note="L7 copies 2 mer ac 97% conserved"
13684. .13715
repeat_region /note="8 copies 4 mer acac 100% conserved"
13752. .13779
repeat_region /note="L4 copies 2 mer ca 96% conserved"
13784. .14048
repeat_region /note="MLTIC repeat: matches 189. .465 of consensus"
14716. .15018
repeat_region /note="AluSc repeat: matches 1. .303 of consensus"
15567. .15842
repeat_region /note="AluJo repeat: matches 3. .305 of consensus"
16752. .16916
repeat_region /note="MIR repeat: matches 63. .227 of consensus"
17700. .17838
repeat_region /note="AluJo/FRAM repeat: matches 163. .301 of consensus"
18111. .18314
repeat_region /note="MLTIA2 repeat: matches 157. .359 of consensus"
18317. .18428
repeat_region /note="28 copies 4 mer ctcc 71% conserved"
18319. .18422
repeat_region /note="52 copies 2 mer ct 63% conserved"
18322. .18423
repeat_region /note="L7 copies 6 mer tctctc 63% conserved"
18510. .18797
repeat_region /note="AluSg repeat: matches 1. .302 of consensus"
18800. .18956
repeat_region /note="MLTIA2 repeat: matches 1. .171 of consensus"
19392. .19704
repeat_region /note="AluJo repeat: matches 1. .312 of consensus"
19780. .20069
repeat_region /note="AluY repeat: matches 1. .290 of consensus"
22285. .22456
repeat_region /note="MIR repeat: matches 23. .220 of consensus"
22457. .22760
repeat_region /note="AluSg repeat: matches 1. .304 of consensus"
22761. .22773
repeat_region /note="MIR repeat: matches 220. .230 of consensus"
26734. .26961
repeat_region /note="L1MB2 repeat: matches 5944. .6170 of consensus"
28074. .28174
repeat_region /note="MER44C repeat: matches 7. .103 of consensus"

repeat_region 28175. .28840
/note="MER72 repeat: matches 5. .668 of consensus"
28841. .29431
repeat_region /note="MER44C repeat: matches 103. .721 of consensus"
29821. .30105
repeat_region /note="AluX repeat: matches 1. .285 of consensus"
30219. .30775
repeat_region /note="L2 repeat: matches 2150. .2749 of consensus"
30924. .31207
repeat_region /note="AluX repeat: matches 3. .294 of consensus"
32053. .32298
repeat_region /note="MIR repeat: matches 1. .257 of consensus"
32498. .32740
repeat_region /note="L1MB3 repeat: matches 5943. .6182 of consensus"
33867. .33920
repeat_region /note="27 copies 2 mer ca 90% conserved"
33871. .33918
repeat_region /note="8 copies 6 mer cacaca 93% conserved"
33878. .33921
repeat_region /note="L11 copies 4 mer acac 97% conserved"
34367. .34420
repeat_region /note="9 copies 6 mer ctcttc 75% conserved"
34458. .34576
repeat_region /note="MIR repeat: matches 94. .219 of consensus"
34914. .35006
repeat_region /note="MIR repeat: matches 150. .242 of consensus"
35044. .35414
repeat_region /note="MLTIB repeat: matches 1. .386 of consensus"
35440. .35480
repeat_region /note="L2 repeat: matches 2423. .2464 of consensus"
35653. .36028
repeat_region /note="THEIB repeat: matches 1. .359 of consensus"
36683. .37012
repeat_region /note="AluX repeat: matches 1. .308 of consensus"
37526. .37818
repeat_region /note="AluJo repeat: matches 1. .290 of consensus"
37957. .38028
repeat_region /note="36 copies 2 mer tg 86% conserved"
37968. .38031
repeat_region /note="L6 copies 4 mer gtgt 90% conserved"
37980. .38027
repeat_region /note="8 copies 6 mer gtgtgt 100% conserved"
38553. .39205
repeat_region /note="L1MB3 repeat: matches 5545. .6183 of consensus"
39561. .39944
repeat_region /note="WSTA repeat: matches 1. .426 of consensus"
40827. .41086
repeat_region /note="AluX repeat: matches 1. .259 of consensus"
41793. .42089
repeat_region /note="AluSg repeat: matches 1. .297 of consensus"
42330. .42393
repeat_region /note="L2 repeat: matches 2627. .2701 of consensus"
42692. .42857
repeat_region /note="L1MC2 repeat: matches 5654. .5816 of consensus"
42858. .43151
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
43152. .43183
repeat_region /note="L1MC2 repeat: matches 5816. .5846 of consensus"
43184. .43479
repeat_region

Query Match 37.9%; Score 107; DB 9; Length 96183;
Best Local Similarity 100.0%; Pred. No. 4.2e-22;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 CAAGAAGCTCTGGGAAGTGTGCTGGATGCAGTTTGGCAACCACTCAACCTCGTGGAA 129
|||||

Db 44203 CAAGAAGCTCTGGGAAGTGTGCTGGATGCAGTTTGGCAACCACTCAACCTCGTGGAA 44262
|||||

QY 130 GTTGACTATTGGCCCTCGAGTTTCTCTGATCACAAAAGATCACGCT 176
|||||

Db 44263 GTTGACTATTGGCCCTCGAGTTTCTCTGATCACAAAAGATCACGCT 44309
|||||

* 100197 100296: gap of 100 bp
* 100297 107015: contig of 6719 bp in length
* 107016 107115: gap of 100 bp
* 107116 114507: contig of 7392 bp in length
* 114508 114607: gap of 100 bp
* 114608 122671: contig of 8064 bp in length
* 122672 122771: gap of 100 bp
* 122772 128052: contig of 5281 bp in length
* 128053 128152: gap of 100 bp
* 128153 138583: contig of 10431 bp in length
* 138584 138683: gap of 100 bp
* 138684 150248: contig of 11565 bp in length
* 150249 150348: gap of 100 bp
* 150349 152053: contig of 1705 bp in length.

FEATURES

source
1. .152053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature
1. .784
/note="assembly_fragment
vector_side:left"
885. .1956
/note="assembly_fragment"
misc_feature
2057. .3113
/note="assembly_fragment"
misc_feature
3214. .4817
/note="assembly_fragment"
misc_feature
4918. .6582
/note="assembly_fragment"
misc_feature
6683. .8059
/note="assembly_fragment"
misc_feature
8160. .22369
/note="assembly_fragment"
misc_feature
22470. .24974
/note="assembly_fragment"
misc_feature
25075. .27206
/note="assembly_fragment"
misc_feature
27307. .29420
/note="assembly_fragment"
misc_feature
29521. .32000
/note="assembly_fragment"
misc_feature
32101. .36330
/note="assembly_fragment"
misc_feature
36431. .39067
/note="assembly_fragment"
misc_feature
39168. .42564
/note="assembly_fragment"
misc_feature
42665. .46664
/note="assembly_fragment"
misc_feature
46765. .51047
/note="assembly_fragment"

Query Match 37.9%; Score 107; DB 2; Length 152053;
Best Local Similarity 100.0%; Pred. No. 4,7e-22;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 CAAAGAGCTCCTGGAGGTCGCTGGATGCAGTTTGCACACCTCAACCTCGTGGAA 129

|||||
Db 31950 CAAAGAGCTCCTGGAGGTCGCTGGATGCAGTTTGCACACCTCAACCTCGTGGAA 31891

QY 130 GGTGACTATTTGGCTCGAGTTCTGTGATCACAAAAGATCAGGT 176

|||||
Db 31890 GGTGACTATTTGGCTCGAGTTCTGTGATCACAAAAGATCAGGT 31844

RESULT 4

BC021301

LOCUS

BC021301

2605 bp

mRNA

linear

PRI 22-JAN-2002

DEFINITION

IMAGE:5013180, mRNA, complete cds.

ACCESSION BC021301

VERSION BC021301.1

GI:18204274

KEYWORDS MGC.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2605)

Strausberg, R.

Direct Submission

Submitted (14-JAN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgabps@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: http://www.nisc.nih.gov/

Contact: nisc_mgc@nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAL Plate: 39 Row: 1 Column: 18

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 7662309.

Location/Qualifiers

1. .2605

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:29593 IMAGE:5013180"

/tissue_type="Brain, astrocytoma, grade IV"

/clone_lib="NIH_MGC_98"

/lab_host="DH10B"

/note="Vector: pOT7"

126. .2069

/codon_start=1

/product="Similar to KIAA0793 gene product"

/protein_id="AAH21301.1"

/db_xref="GI:18204275"

/translation="MGEIEGTQVLTQAGRLGCAQTPVGSTLEPGQTLPRMQEHL

HLRVKLLDNTMEIFDIEPCDQVLLTQVKNLNVCEDFGMEFQNTQSYWTLPEPM

KPIIRRRKNVLRVAVFPDQGLQEEVRYLQALQRLLERLCADTFTA

ALLTSHLQSEIGDYDREHLKKNVYLPQGHCKLEKILFEHQHVGQTPAESDFQ

VLEIARKLMYGIREFHMASDREGTKIQLAVSHMGVLFVCGTTKINTFNWVKVLSFK

RRKRLKLFHVEVHPYQDTLEFLGSRDECKNFKICVETHTFFLLDQDQPKAKAVF

FRGSGFRYSGRGTQKQVDFKDSGNKRIPYERRHSHKTHTSVRLTADLPKQISFPE

GRTPTASQAPLGPALQPGSLSTKSPQSPSSRSPLSLSPAFVPLGPAGSGSPL

LSPVLSDAGGACMDCDEPRHKVPADAEVIVKEILATERTYLLKLEVTYVFRNAV

KEDAPATLTLLEFNSIDPIEFHKGFLREVEQRLALWEGPSKAHTKGSHQRIGDILL

RNRQLKVFQLHFGHVAGVTKE"

BASE COUNT 675 a 697 c 662 g 571 t

ORIGIN

Query Match

Best Local Similarity

Matches 165; Conservative

34.7%; Score 97.8; DB 9; Length 2605;

59.6%; Pred. No. 1.5e-19;

0; Mismatches 112; Indels 0; Gaps 0;

VERHYARDQNNELIGVMSGGLIIVKYNRMVMTFLWLKVIKISFKCKOFFIOLRKEL
HESRETLGFMVYVYRACKTLWACVEHHTFFRLGPLPPKNNFFAHFYTLGSKFRYC
GRTEVQVQYKGNKAMDVFRARSSKGLVHVSQDQVVRNKLSDRDLTQSLPRLS
PGTPNHRNSFTQEARVFRPSSVGHVLDHVSQDQVVRNKLSDRDLTQSLPRLS
SPQEPNEDQDQALPKQSKNSWQIHFESNQDQVVRNKLSDRDLTQSLPRLS
GIPHNILVLMKMPDENRGFPNKGQDQVVRNKLSDRDLTQSLPRLS
VLINGRDIAHTDQVLFIFKASCEHRSGLVLRNNAVIVSRVAPGTPADLCVPRNEGDOV
EKAPLDSVHDDHSLRESMTQALAEGLITGVLAQFQDLYRKFKGFMSCAKLPONISK
NRYRDISPYDQTRVLKNGEDYINANYINNEIPSSSINOYIACQGPLPHTCKDFWQM
IWQGSVMYVMTLTOVERGRVKCHOWPESESSYGYOATCSHSEGNPAYIFRKMWT
LIQKRNESQLTQIQYTAMPDHGVDPDSDSDFLDFVCHVDQDQAGKEEPIIIVCSAGI
GRGVLLTMTAMCLIECPNQPVLDIRVTRDQRAMMIQTPSQYRVCBAILKYEE
GFVKPLTSSNK"

BASE COORD 920 a 587 c 592 g 783 t

Query Match 25.48; Score 71.6; DB 10; Length 2882;
Best Local Similarity 57.38; Pred. No. 2.8e-11;
Matches 150; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY 24 CAAMATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCACAAAGAGCTCTGG 83
DB 108 CAACATCTTCTTCGGATACACTGTACAGCTTCAGAGTTACAAACATGATCAGG 167
QY 84 GAAGTGTGCTGGATGAGTTGCAACACCTCAACCTCGTGAAGTGACTATTTGG 143
DB 168 GCAAGTCTGTGGATAGTCTTCAAGCATCTTGATTGACTGAGCGAGACTATTTGG 227
QY 144 CCTCAGTTTCT---GATCACAACAAATCAGGTGCGTGGATCCCTTAAACCCAT 200
DB 228 TTACAGTTGGTGGATGAGTTCACAGATACCAAGGTGGTGGATCCCAACCAACCAAT 287
QY 201 TGTGAACAGATTGAAGCCCAAGCAGCTTGTGTTAAGTTTGGTGAATTTCTTCC 260
DB 288 AAGGAGCAGTAAGAGAGGATCACCITACAAATTGACITTAGAGTCAATTTCTTGT 347
QY 261 GCCTGACCAACACAACTCCAA 282
DB 348 AAGTGACCCCAAGGTTACAA 369

RESULT 10 ALI36300 170916 bp DNA linear PRI 22-JUN-2001
LOCUS Human DNA sequence from clone RP11-10G5 on chromosome 13 Contains a
zinc finger gene similar to ZNF183, the start of the FARP1 (FERM,
RHOGEF (ARHGEF) and pleckstrin domain protein 1
(chondrocyte-derived)) gene, ESTs, STSS, GSSs and CpG islands,
complete sequence.
ACCESSION ALI36300.22 GI:11125392
VERSION HTG; FARP1; FERM; RHOGEF; zinc finger.
KEYWORDS human.
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170916)
AUTHORS Smith,M.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT On Nov 8, 2000 this sequence version replaced gi:11061750.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
This sequence is the entire insert of clone RP11-10G5 The true left
end of clone RP11-573N10 is at 109104 in this sequence. The true
right end of clone RP11-72J7 is at 95284 in this sequence. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-10G5 is from the
library RPCI-11.1 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
source 1. .170916
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-10G5"
/clone_lib="RPCI-11.1"
384. .873
/note="MER31B repeat: matches 1. .481 of consensus"
repeat_region 1374. .1407
/note="17 copies 2 mer tt 85% conserved"
repeat_region 5470. .5533
/note="32 copies 2 mer tt 70% conserved"
misc_feature 6352. .6714
/note="match: GSS: Em:AQ698072"
repeat_region 8498. .8665
/note="TIGGER1 repeat: matches 2080. .2238 of consensus"
repeat_region 8670. .8888
/note="MER75B repeat: matches 18. .239 of consensus"
repeat_region 13335. .13686
/note="176 copies 2 mer tg 69% conserved"
repeat_region 13996. .14238
/note="L1PB3 repeat: matches 5894. .6150 of consensus"
repeat_region 15610. .15639
/note="15 copies 2 mer ta 86% conserved"
misc_feature complement(17881. .18076)
/note="match: GSS: Em:AQ76608"
repeat_region 18339. .18804
/note="LTR7 repeat: matches 1. .448 of consensus"
repeat_region 19891. .20105
/note="MER57-internal repeat: matches 7171. .7387 of
consensus"
20131
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 260bp by restriction digest data."
repeat_region 29106. .29251
/note="LTR41 repeat: matches 2. .160 of consensus"
repeat_region 31944. .32589
/note="LTR12 repeat: matches 1. .671 of consensus"
repeat_region 32590. .33323
/note="PTB5 repeat: matches 743. .1478 of consensus"
repeat_region 38726. .38851
/note="63 copies 2 mer ca 88% conserved"
misc_feature 41206. .41622
/note="match: GSS: Em:AQ184690"
misc_feature 42604. .43109
/note="match: GSS: Em:AQ573113"
repeat_region 43382. .43429
/note="24 copies 2 mer ta 93% conserved"
misc_feature complement(46659. .47093)
/note="match: GSS: Em:AQ805998"
repeat_region 46717. .46754
/note="19 copies 2 mer gt 100% conserved"
repeat_region 53170. .53274

repeat_region /note="MER34 repeat: matches 8. .115 of consensus"
53278. .53412
/note="MER74B repeat: matches 50. .180 of consensus"
55996. .56438
/note="MER83 repeat: matches 1. .448 of consensus"
59536. .60314
/note="MER21B repeat: matches 63. .794 of consensus"
60344. .60392
/note="MER21B repeat: matches 741. .790 of consensus"
60825. .61336
/note="MER74B repeat: matches 61. .621 of consensus"
61616. .61665
/note="MER61B repeat: matches 377. .425 of consensus"
61667. .62339
/note="LTR9 repeat: matches 1. .625 of consensus"
62342. .62683
/note="MER61C repeat: matches 1. .384 of consensus"
63996. .64135
/note="MLT2E repeat: matches 1. .149 of consensus"
64486. .64768
/note="MLT2E repeat: matches 149. .395 of consensus"
65150. .65369
/note="L1MA5A repeat: matches 6053. .6291 of consensus"
65371. .66037
/note="LTR12 repeat: matches 1. .647 of consensus"
66038. .66088
/note="L1MA5A repeat: matches 6037. .6091 of consensus"
66076. .66299
/note="L1MA5A repeat: matches 5827. .6047 of consensus"
67351. .67427
/note="MLT2E repeat: matches 173. .250 of consensus"
67428. .67487
/note="30 copies 2 mer ag 88% conserved"
67492. .67660
/note="MLT1A2 repeat: matches 5. .174 of consensus"
68268. .68810
/note="MER68A repeat: matches 1. .561 of consensus"
complement(73998. .74451)
/note="match: GSS: Em:AQ695161"
74143. .74773
/note="LTR39 repeat: matches 172. .794 of consensus"
75063. .75233
/note="LTR39 repeat: matches 1. .172 of consensus"
79181. .79291
/note="MLT1A2 repeat: matches 251. .374 of consensus"
79699. .80013
/note="MLT1A2 repeat: matches 1. .251 of consensus"
80095. .80463
/note="THE1C repeat: matches 1. .369 of consensus"
83250. .83460
/note="rigger3b repeat: matches 1. .210 of consensus"
83757. .84298
/note="rigger3b repeat: matches 210. .750 of consensus"
84581. .84607
/note="rigger3b repeat: matches 750. .775 of consensus"
84608. .84963
/note="THE1C repeat: matches 4. .371 of consensus"
84964. .85004
/note="rigger3b repeat: matches 775. .825 of consensus"
85677. .86088
/note="rigger3b repeat: matches 823. .1231 of consensus"
complement(89463. .90009)
/note="match: GSS: Em:AQ013482"
complement(89490. .90015)
/note="match: GSS: Em:AQ383425"
complement(89526. .90008)
/note="match: GSS: Em:B82154"
complement(93764. .94093)
/note="match: GSS: Em:AQ594661"
98631. .100473
/note="CpG island"
evidence-not_experimental
99666. .99802

misc_feature /note="match: GSS: Em:AQ938343"
99694. .99750
/note="single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
100889. .100712
/note="12 copies 2 mer tg 100% conserved"
101250. .101669
/note="match: GSS: Em:AQ008502"
101284. .101479
/note="match: GSS: Em:AQ000309"
105297. .105359
/note="L1MEC repeat: matches 1638. .1710 of consensus"
108760. .109211
/note="match: GSS: Em:B66665"
108771. .109149
/note="match: GSS: Em:AQ076004"
108818. .109537
/note="match: GSS: Em:AQ307265"
109801. .110016
/note="108 copies 2 mer at 58% conserved"
117543. .118068
/note="match: GSS: Em:AQ678189"
118206. .118247

Query Match . 25.1%; Score 70.8; DB 9; Length 170916;
Best Local Similarity 97.3%; Pred. No. 1.2e-10;
Matches 72; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 COTTCAGGAAACTCGTGCATCAATAATCCAGATGCTGGATGACACCCAGAGGCATT 60
|||||
Db 169831 COTTCAGGAAACTCGTGCATCAATAATCCAGATGCTGGATGACACCCAGAGGCATT 169890
|||||

QY 61 GAAGTTCACCAAG 74
|||||

Db 169891 GAAGTTCACGTAAG 169904

RESULT 11
AB032366 Locus AB032366 3451 bp mRNA linear ROD 23-MAY-2000
DEFINITION Mus musculus Ehm2 mRNA, complete cds.
ACCESSION AB032366
VERSION AB032366.1 GI:8051681
KEYWORDS EHM2.
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Hashimoto,Y., Shindo-Okada,N., Tani,M., Takeuchi,K., Toma,H. and
Yokota,J.
TITLE Identification of genes differentially expressed in association
with metastatic potential of K-1735 murine melanoma by messenger
RNA differential display
Cancer Res. 56 (22), 5266-5271 (1996)
JOURNAL 97069887
MEDLINE 2 (sites)
REFERENCE 2 (sites)
AUTHORS Shimizu,K., Nagamachi,Y., Tani,M., Kimura,K., Shiroyishi,T.,
Wakana,S. and Yokota,J.
TITLE Molecular cloning of a novel NF2/ERM/4.1 superfamily gene, ehm2,
that is expressed in high-metastatic K1735 murine melanoma cells
Genomics 65 (2), 113-120 (2000)
JOURNAL 20247250
MEDLINE 3 (bases 1 to 3451)
REFERENCE 3 (bases 1 to 3451)
AUTHORS Yokota,J., Shimizu,K. and Nagamachi,Y.
TITLE Direct Submission
Submitted (04-SEP-1999) Jun Yokota, National Cancer Center Research
Institute, Biology Division; Tsukiji 5-chome 1-1, Chuo-ku, Tokyo
104-0045, Japan {E-mail:jyokota@ncc.res.ncc.go.jp,
Tel:81-3-3547-5272, Fax:81-3-3542-0807}

FEATURES
Location/Qualifiers
1. .3451
source

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
1..3451
/gene="Ehm2"
499..2082
/gene="Ehm2"
/codon_start=1
/product="Ehm2"
/protein_id="BAA96078.1"
/db_xref="GI:8051682"
/translation="MLRLRFTGRRSQRYARGAAGGAGLGRDGRGPRGPAA
ASSVLPAGGSVFPAGGGPLTGGAAVHISASGAATKATLYLIDTEFVSDLP
KHAQQLDFDQVHLDELVDYFGLQFLDQAQVTHLDHAKPIKKQKMGVPAVALHF
RVKYSSEPNLREFFRYLFVLOLRHDLISGLKPCPYETAVALAALCLQALGECGL
PEHPULVEERFFIPNQTEAMEFDIFQWKEKYSQAQAEISLNAKWLKEMGVDM
HVGROCEYSLGLTPIGLIFEGANKIGLFWPKTKIMDFKKSKLTLVVVEDDDGG
REQHTFVRLDSARTKHLWKCAVEHHAFFRLRTPSNKSARSDFIRLGRFRFSR
TEYQATGSLRLRTSTFRKPSKRYPSRRHSTFRASNPITAAQLCSKANPEVHNYQO
YHPDVHPSQPRWRPSPHNSHISCKQKPCFQDDRPHWKTSSASGDSGHFDYVQDQO
RNLGGAYSVTYRDKLMTAL"
BASE COUNT 764 a 958 c 967 g 762 t
ORIGIN
```

```
Query Match 25.0%; Score 70.4; DB 10; Length 3451;
Best Local Similarity 55.2%; Pred. No. 6.9e-11;
Matches 137; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 35 TCCTGGATGACACCCAGGAGGATTTGAAGTTCACAAAGAGCTCTGGGAGGTGCTGC 94
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 TCCTGACGGAACCGAAGTGAGCGTGGACCTCGCGAAACATGCCAAAGCCAGGATTTGT 829

QY 95 TGGATGCACTTTGCAACACCTCAACCTCGTGAAGGTGACTATTTCGCCCTCGAGTTTC 154
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 TGACAGATTTGTACCACTTGGACCTGGTGGTGAACAGATTTATTTCGTCGAGTTCC 889

QY 155 CTGATCACAAAAGATGACGGTGTGGCTGATCTCTCTAAAACCCATTTGGAACAGATTA 214
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 TCGACTCTGCCAGGTGACGCACTGCTGGATCAGCGAAGCCATTAAGAAGAGATGA 949

QY 215 GAAGGCCAAGACAGTTGTTGTTAGTTGTTGGTGAATTCCTTCGCCCTGACCAACAC 274
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 AAGTGGGACCTGCTATGCTTGGCACTTTGAGTCAAAATACTATTCTTCAGAACCAACA 1009

QY 275 AACTCAA 282
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1010 ACCTCGA 1017
```

```
RESULT 12
LOCUS BC010674 3619 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, protein tyrosine phosphatase, non-receptor type 4
(megakaryocyte), clone MGC:9204 IMAGE:3853914, mRNA, complete cds.
ACCESSION BC010674
VERSION BC010674.1 GI:14715026
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3619)
Strausberg, R.
Direct Submission
Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalob@bcm.tmc.edu.
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: b Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 190747.

FEATURES

Location/Qualifiers
1..3619
/organism="Homo sapiens"
/db_xref="locusID:5775"
/db_xref="taxon:9606"
/clone="MGC:9204 IMAGE:3853914"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
428..3208
/codon_start=1
/product="protein tyrosine phosphatase, non-receptor type
4 (megakaryocyte)"
/protein_id="AAH10674.1"
/db_xref="GI:14715027"
/translation="MTRSRFLPAGRTYNNVRASELARDROHTEVVVNCILLDNTVOAFK
VNHDOGVLLDVVFKHLDLTEQDYFGLQADLSDTNPRLDNPKNPKIRKOLKRGSPYS
LNFVKFVSDPNKQLBEYRYFLOIQLDILGRPLPCPSNTAALLASFAVQSELGD
YDSENLGSLYSDFIPNQPDFEKEIAKLHQHISPAEAEFLNTANTLDELVG
VEFYARDQSNNEIMIGMSGILIKYKRVMTFPWLKPKVIFKFKQFFIOLRKL
HESRETLGFMVNRACKNLWAKCAVEHHTFFRLDRPLPKPKNFAHFYFTLGSFRVC
GRTEVQSVQYKKEKANDRVFARSPKPLAKMLDWEVSNISLDRLETSQSPRS
PGTPTNHRNSTFTQETRLRPSVGHVLDHVMVHTSPSEFVNQRSPSTQANSVLIES
SPSETPGDKGPALPPKKNQNLHYSHQQDLLESHNETFTDIPSPKEPTNG
VLINGRDLVLRMKPDQVDFNFYKGGYDQKMPVIVSRVAPGTGTPADLCVRLNLEGDQV
VLINGRDIAEHTHDQVYLFKACSERHSGELMLLRVNAVYDVVVEERKLENEPQYIP
GLPADSVHODDHSRESMLQAEGLITVTFQDOLYRKPKGMWMSKALPQNIK
NRYRDISPYDATRVILKNEDYINANYINMEIPSSIIINOYIACQGPLPHTCTDFWOM
TWEQSSWVWMLTTOVERGRVKCHQYHPEPTGSSSYGCVQVCHSEGNATYIFRKMT
LFNQEKNESRPLTQIAMPDGVDDSDFLDFVCHVRNKRAGKEPVPVHCSAGI
GRGTGLITMETMALIECNOPVYPLDIVRTMRDORAMMIOTPSQYRFVCEAILKYEE
GFVKPLTSTNK"

```
BASE COUNT 1116 a 762 c 789 g 952 t
ORIGIN

Query Match 24.3%; Score 68.4; DB 9; Length 3619;
Best Local Similarity 56.5%; Pred. No. 3e-10;
Matches 148; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 24 CAAATTCAGATGCTGATGACACCCAGGAGCATTTTGAAGTTCACAAAGAGCTCTGG 83
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CAACATCTCTTCGATACACCTGTACAGCTTCAAGTCAATAACATGATCAGG 579

QY 84 GAAGTGTGCTGGATGAGTTTGAACCACTCAACCTCGTGAAGGTGACTATTTTGG 143
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 GCAAGTCTTGTGGATGCTGCTTCAAGCATCTAGATTTGACTGAGCAGGACTATTTGG 639

QY 144 CCTCAGTTTCTGATCA---CAAAAAGATCACGGTGTGGTGGATCTCCCTAAACCAT 200
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 TTTACAGTTGCTGATGATTCACAGATAACCAAGGTGGCTGGATCCAAACAAACCAAT 699

QY 201 TGTGAACAGATTAGAAGCCAAAGCACGCTTCTTAAAGTTTGTGTAATTTCTTCC 260
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 AAGGAGCAGCTAAGAGAGGATCTCCTTACAGTTTGACTTTAGATCAAAATTTTGT 759

QY 261 GCCTGACCAACACAACTCCAA 282
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Df 760 AAGTACCCCAACAAGTTACAA 781

RESULT 13
HUMPTYPH
LOCUS HUMPTYPH 3643 bp mRNA linear PRI 27-APR-1993
DEFINITION Human protein-tyrosine phosphatase mRNA, complete cds.
ACCESSION M68941
VERSION M68941.1 GI:190747
KEYWORDS protein-tyrosine phosphatase.
SOURCE Homo sapiens (library: cDNA, Meg-01, HUVEC) CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Gu,M.X., York,J.D., Marshawsky,I. and Majerus,P.W.
TITLE Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to cytoskeletal protein 4.1
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (13), 5867-5871 (1991)
MEDLINE 91288564
PUBMED 1648233
FEATURES
source Location/Qualifiers
1..3643 /organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Meg-01, human platelets"
/tissue_lib="cDNA, Meg-01, HUVEC"
<1..3643
/product="protein-tyrosine phosphatase"
1..771
/note="contains short ORF's, GC rich stem loop structure with 75% homology to CABL protooncogene 5'UTR; putative"
632..673
/note="GC rich stem loop structure"
772..3552
/codon_start=1
/product="protein-tyrosine phosphatase"
/protein_id="AAA36530.1"
/db_xref="GI:190748"
/translations="MFSRFLPAGRTYNNVRASLEARDROHTVEVVCNILLIDNTVQAFK VNKHDGQVLVDVVFHLLDTEQDFGLQADDSSTONRWLDNPKPIRKOLKRGSPYS LNPFRFVSFDKNLGEQETRYQFIQIKODITGLRPLCPSTAALLAGFAVGSELGD YNQSEALSYSDSPYIPNQPDFEKIAKHQQHLGLSPAEAFNYLNARTLELYG VESFYARDQSNMIMIGVMSSGGILIIYNRMVMTFWLTKVISFKCKOFFIQLRELK HEPTRLGFNMVNMYRNACKLKACVEHTEFFRLDRPLLPKKNEFAHYFTLGSKFRC GRTEVOYQYGEKANCDRVFARPSKPLARKLMDEWVSRNSISDDRLETQSLSPRS PGTPPNHRNSTFTCGEPTRSPSVSHLVDMVHTSPSEVFVNQRSPSQTSQANSIVLES PSQETPGDGKPLQSPKSKNSNQIHYSHQDLSEHINFETDIPSPKEPTNG GPHDNLIATIRMPDPBNGRFKFNKGKYDQMPEIVSRVAPGTADLCVPRLMGEQV VLINGDIAEHTHDQVLFITKASCRHSGLMLLRPNAYDVVEEKLNEPDFOVIP EKAPLDSVHDDHSLRESMTQAEGLITGVLTQFDOLYVKKPGMTMSCAKLPQNISK NRYRDISPVHDKMGEDINYANINWEIPSSSIINQYIACQGPHPTCTFNMWM TPDQSGMVMLTQVERGVKQHWPEPTGSYYCYQVTCHSEBGNATYIFRRKT LFNQKNERSEPQTIOQYIAMPDHGVPDSDSDFDYFCHVRNKRAGKEEPPVYHCASI GRTVGLITMETAMCLIECNQPYPPLDIVRTMRDQRAMMIQTGPSQYRFVCEAILKVVEE GFVKPLTTSTNK"

BASE COUNT 1038 a 815 C 863 g 927 t

ORIGIN

Query Match 24.3%; Score 68.4; DB 9; Length 3643;
Best Local Similarity 56.5%; Pred. No. 3e-10;
Matches 148; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 24 CAATAATCCAGATGCTGGATGACACCAGAGGCCATTGAAGTTCCACAAAGAGTCCTGG 83
||| |||| | ||||| |||| | ||||| |||| | ||||| |||| | ||||| ||||
Db 864 CAACATCCTTCTTGATGATACACTGTACAGAGCTTTCAAAGTCAATAAATGATGATCAGG 923
||| |||| | ||||| |||| | ||||| |||| | ||||| |||| | ||||| ||||
QY 84 GAAGGTGCTGCTGGATGACAGTTGCAACCACCTCGTGAAGGTGACTATTITGG 143
||| |||| | ||||| |||| | ||||| |||| | ||||| |||| | ||||| ||||
Db 924 GCAAGTCTTGATGCTGCTTCAAGCATCTATAGATTGACTGAGCAGGACTATTITGG 983
||| |||| | ||||| |||| | ||||| |||| | ||||| |||| | ||||| ||||
QY 144 CCTCGAGTTTTCTGATCA---CAAAAAGATCACGGTGTGGCTGGATCTCTCTAAACCCAT 200

Db	984	TTTACAGTTGCTGATGATTCACAGATAACCCAGGTGGCTGGATCAAAACAACCAAT	1043
Qy	201	TGTTGAACAGATTAGAGGCCAAAGCAGCTGTGCTTAAGTTTGTGGTGAATTCCTTTCC	260
Db	1044	AAGGAAGCAGCTAAGAGAGAGATCCTCTACAGTTTGAAGTTTGAAGTCAAAATTTTGT	1103
Qy	261	GCCTGACCACCAACAATCCAA	282
Db	1104	AAGTGACCCCAACAAGTTACAA	1125
RESULT 14			
LOCUS	AC022669	152053 bp	DNA linear
DEFINITION	Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT		
ACCESSION	AC022669		
VERSION	AC022669.4	GI:9966291	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 152053)		
TITLE	Barren,B., Linton,L., Nusbaum,C. and Lander,E.		
JOURNAL	Homo sapiens chromosome 13, clone RP11-96B23		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 152053)		
	Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burckett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 152053)		
AUTHORS	Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacLean,C., MacDonald,P., Marquis,N., Matthews,K., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Rette,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7229803.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4938
Center clone name: 96_B_23
----- Summary Statistics
Sequencing vector: M13; M7815; 93% of reads
Sequencing vector: Plasmid; n/a; %0.f% of reads
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 136451 bases at least Q40
Consensus quality: 143612 bases at least Q30
Consensus quality: 146365 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 148853; sum-of-contigs
Quality coverage: 3.4 in Q20 bases; agarose-fp
Quality cov.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
785 884: contig of 784 bp in length
885 1956: contig of 1072 bp in length
1957 2056: gap of 100 bp
2057 3113: contig of 1057 bp in length
3114 3213: gap of 100 bp
3214 4817: contig of 1604 bp in length
4818 4917: gap of 100 bp
4918 6582: contig of 1665 bp in length
6583 8059: contig of 1377 bp in length
8060 8159: gap of 100 bp
8160 22369: contig of 14210 bp in length
22370 22469: gap of 100 bp
22470 24974: contig of 2505 bp in length
24975 25074: gap of 100 bp
25075 27206: contig of 2132 bp in length
27207 27306: gap of 100 bp
27307 29420: contig of 2114 bp in length
29421 29520: gap of 100 bp
29521 32000: contig of 2480 bp in length
32001 32100: gap of 100 bp
32101 36330: contig of 4230 bp in length
36331 36430: gap of 100 bp
36431 39067: contig of 2637 bp in length
39068 39167: gap of 100 bp
39168 42564: contig of 3397 bp in length
42565 42664: gap of 100 bp
42665 46664: contig of 4000 bp in length
46665 46764: gap of 100 bp
46765 51047: contig of 4283 bp in length
51048 51147: gap of 100 bp
51148 55269: contig of 4122 bp in length
55270 55369: gap of 100 bp
55370 58831: contig of 3462 bp in length
58832 58931: gap of 100 bp
58932 64077: contig of 5146 bp in length
64078 64177: gap of 100 bp
64178 68808: contig of 4631 bp in length

* 68809 68908: gap of 100 bp
68909 74170: contig of 5262 bp in length
74171 74270: gap of 100 bp
74271 79296: contig of 5026 bp in length
79297 79396: gap of 100 bp
79397 84724: contig of 5328 bp in length
84725 84824: gap of 100 bp
84825 89713: contig of 4889 bp in length
89714 89813: gap of 100 bp
89814 94721: contig of 4908 bp in length
94722 94821: gap of 100 bp
94822 100196: contig of 5375 bp in length
100197 100296: gap of 100 bp
100297 107015: contig of 6719 bp in length
107016 107115: gap of 100 bp
107116 114507: contig of 7392 bp in length
114508 114607: gap of 100 bp
114608 122671: contig of 8064 bp in length
122672 122771: gap of 100 bp
122772 128052: contig of 5281 bp in length
128053 128152: gap of 100 bp
128153 138583: contig of 10431 bp in length
138584 138683: gap of 100 bp
138684 150248: contig of 11565 bp in length
150249 150348: gap of 100 bp
150349 152053: contig of 1705 bp in length.

FEATURES

source
1..152053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13"
/clone="RP11-96B23"
/clone_lib="RPC1-11 Human Male BAC"
1..784
/note="assembly_fragment"
clone_end:SP6
vector_side:left
885..1956
/note="assembly_fragment"
2057..3113
/note="assembly_fragment"
3214..4817
/note="assembly_fragment"
4918..6582
/note="assembly_fragment"
6683..8059
/note="assembly_fragment"
8160..22369
/note="assembly_fragment"
22470..24974
/note="assembly_fragment"
25075..27206
/note="assembly_fragment"
27307..29420
/note="assembly_fragment"
29521..32000
/note="assembly_fragment"
32101..36330
/note="assembly_fragment"
36431..39067
/note="assembly_fragment"
39168..42564
/note="assembly_fragment"
42665..46664
/note="assembly_fragment"
46765..51047
/note="assembly_fragment"

Query Match 23.2%; Score 65.4; DB 2; Length 152053;
Best local Similarity 98.5%; Pred. No. 6e-09; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 1;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 14:48:04 ; Search time 44.5773 Seconds
(without alignments)
14246.337 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_432

Perfect score: 282

Sequence: 1 ccttcaggaaactcgtgtc.....ctgaccacacacaaactcaa 282

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	100.0	3187	23 AAS64760	DNA encoding novel
2	282	100.0	3442	20 AAX79183	Human chondrocyte-
3	282	100.0	4687	21 AAA08582	Human cytoskeleton
c 4	138	48.9	241	AAA42150	Human secreted exp
5	68.4	24.3	3643	23 ABK84074	Human cDNA differe
6	65.2	23.1	659	22 AAH33845	Human colon cancer
7	65.2	23.1	1234	23 AAS73863	DNA encoding novel
8	65.2	23.1	2872	23 AAS08887	Human protein tyro
9	65.2	23.1	3313	22 ABLN93400	Human gene GS91552

10	61.4	21.8	14917	24 AAD32338	Human lung specifi
11	60.6	21.5	3106	23 ABN93399	Human gene GS91552
12	57.8	20.5	621	24 ABQ72602	Human MDDT encodin
13	57.8	20.5	2846	23 ABL60869	Human cytoskeleton
14	57.8	20.5	4922	23 ABLV30214	Human prostate exp
15	56.8	20.1	6263	24 ABK70019	Human cDNA for the
16	56.6	20.1	3031	23 ABV22627	Human prostate exp
17	56.6	20.1	3031	23 ABV28450	Human prostate exp
18	54.6	19.4	3984	18 AAT58627	Protein tyrosine p
19	54.6	19.4	3984	20 AAX78463	Human PTPH1 CDNA.
20	54.6	19.4	3984	24 AAT72918	PTPH1 CDNA. Homo
21	54	19.1	550	22 AAH06244	Human cDNA clone (
22	54	19.1	2067	24 ABN85314	Human cytoskeleton
23	54	19.1	3166	22 AAH17787	Human cDNA sequenc
24	54	19.1	4356	23 ABL29969	Drosophila melanog
25	52.8	18.7	556	22 AAH99189	Human protein enco
26	47.6	16.9	3620	21 AAA08581	Human cytoskeleton
27	46.6	16.5	2337	24 ABQ73378	Human red blood ce
28	45.8	16.2	1626	22 AAI59140	Human polynucleoti
29	45.8	16.2	1685	24 ABL50340	Human cancer cell
30	45.8	16.2	2556	22 AAI60926	Human polynucleoti
31	45	16.0	1220	23 ABL02335	Drosophila melanog
32	44.6	15.8	347	21 AAC02360	Human secreted pro
33	42.6	15.1	8105	23 ABL23968	Drosophila melanog
34	41.8	14.8	427	21 AAC01157	Human secreted pro
35	39.2	13.9	650	22 AAS31134	Human diagnostic a
36	39.2	13.9	2066	24 ABL60867	Human cytoskeleton
c 37	38.6	13.7	414	24 ABL66151	Lung cancer relate
38	37.2	13.2	336	21 AAC01319	Human secreted pro
39	37.2	13.2	512	24 ABK70156	Human lung cancer
40	37.2	13.2	562	21 AAA69837	Human ovarian carc
41	37.2	13.2	562	24 ABN72731	Ovarian carcinoma
42	37.2	13.2	2930	24 ABQ88181	Human osteoblast d
43	37.2	13.2	2930	24 ABK70285	Human lung cancer
44	37.2	13.2	3044	24 ABQ88180	Human osteoblast d
45	37.2	13.2	3044	24 ABK84552	Human cDNA differe

ALIGNMENTS

RESULT 1

AAS64760

ID AAS64760 standard; cDNA; 3187 BP.

XX AC

AC AAS64760;

XX DT

13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #564.

XX KW

Human; chromosome mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS

Homo sapiens.

XX PN

WO200175067-A2.

XX PD

11-OCT-2001.

XX PF

30-MAR-2001; 2001WO-US08631.

XX PR

31-MAR-2000; 2000US-0540217.

XX PR

23-AUG-2000; 2000US-0649167.

XX PA

(HYSE-) HYSEQ INC.

XX PI

Drmanac RT, Liu C, Tang YT;

XX DR

WPI; 2001-639362/73.

XX DR

P-PSDB; ABG00573.

XX PT

New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 564; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3187 BP; 810 A; 874 C; 876 G; 527 T; 0 other;

Query Match 100.0%; Score 282; DB 23; Length 3187;
Best Local Similarity 100.0%; Pred. No. 3.9e-83;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
|||||
Db 151 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210
QY 61 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 120
|||||
Db 211 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 270
QY 121 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 180
|||||
Db 271 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 330
QY 181 CTGGATCTCTAAACCCATTGTAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 240
|||||
Db 331 CTGGATCTCTAAACCCATTGTAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 390
QY 241 TTTGTGGTGAATTTCTTCCGCGCTGACACACACAACTCCAA 282
|||||
Db 391 TTTGTGGTGAATTTCTTCCGCGCTGACACACACAACTCCAA 432

RESULT 2

AA79183
ID AAX79183 standard; DNA; 3442 BP.

XX AAX79183;

DT 17-AUG-1999 (first entry)

XX Human chondrocyte-derived gene CDEP.

DE Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;
KW drug; ss.

XX Homo sapiens.

OS W0928458-A1.

PN

PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-JP05348.
XX
PR 27-NOV-1997; 97JP-0342060.
XX
PA (CHUS) CHUGAI SEIYAKU KK.

PI Kato Y, Kawamoto T, Koyano Y;
XX
XX WPI; 1999-371117/31.
DR P-PSDB; AAY07482.

XX protein CDEP expressed in differentiated chondrocytes, and gene
PT encoding it

XX Claim 5; Fig 1; 59pp; Japanese.

XX This sequence represents the coding region for a protein (CDEP) expressed
CC in differentiated human foetal chondrocytes, which contains an ezrin-like
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.
CC The nucleic acid or protein can be used in the investigation and
CC treatment of cancers and arthritic diseases (including chronic rheumatoid
CC arthritis), or for screening of candidate anticancer drugs.

XX Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Match 100.0%; Score 282; DB 20; Length 3442;
Best Local Similarity 100.0%; Pred. No. 4.1e-83;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
|||||
Db 151 CCTTCAGGAAACTCGTGTCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 210
QY 61 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 120
|||||
Db 211 GAAGTTCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGCAATTCGCAACACCTCAAC 270
QY 121 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 180
|||||
Db 271 CTCGTGGAAGTGACTATTTTGGCCCTCGAGTTTCTGATCACAAAAGATCACGGTGTGG 330
QY 181 CTGGATCTCTAAACCCATTGTAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 240
|||||
Db 331 CTGGATCTCTAAACCCATTGTAACAGATTAGAGGCCAAGACACGTTGTTGTTAAG 390
QY 241 TTTGTGGTGAATTTCTTCCGCGCTGACACACACAACTCCAA 282
|||||
Db 391 TTTGTGGTGAATTTCTTCCGCGCTGACACACACAACTCCAA 432

RESULT 3

AAA08582
ID AAA08582 standard; DNA; 4687 BP.

XX AAA08582;

DT 19-JUL-2000 (first entry)

XX Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.

KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory; vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 90..3227

FT /*tag= a

FT /product= CYSKP-2

XX

PN WO200017355-A2.
XX
PD 30-MAR-2000.
XX
PF 17-SEP-1999; 99WO-US21565.
XX
XX 18-SEP-1998; 98US-0172226.
PR 27-APR-1999; 99US-0131321.
XX
XX (INCY-) INCYTE PHARM INC.
PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;
XX
XX WPI; 2000-283582/24.
DR P-PSDB; AAY91947.
XX
XX Human cytoskeleton associated proteins, used to treat cell
PT proliferative, autoimmune/inflammatory, vesicle trafficking,
PT neurological, cell motility, reproductive and muscle disorders
XX
XX Claim 9; Page 101-102; 113pp; English.
XX
XX AAA08581-96 encode human cytoskeleton associated proteins 1 to 16
CC (CysRP-1 to CysRP-16) respectively. The sequences can be used to treat
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,
CC vesicle trafficking, neurological, cardiovascular, cell motility,
CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CysRP-2 to CysRP-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CysRP (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
CC cancers, autoimmune/inflammatory disorders such as allergies, anemia,
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CysRP antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CysRP
CC (claimed).
XX
XX Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;
SQ

Query Match 100.0%; Score 282; DB 21; Length 4687;
Best Local Similarity 100.0%; Pred. No. 4.7e-83;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGAAACTCGTGTCCATCAAAATCCAGATGCTGATGACACCCAGAGGCATT 60
Db 192 CCTTCAGAAACTCGTGTCCATCAAAATCCAGATGCTGATGACACCCAGAGGCATT 251
QY 61 GAAGTTCCCAAGAGCTCTCGGAAGGTCTGCTGATGCGATTTGCAACCTCAAC 120
Db 252 GAAGTTCCCAAGAGCTCTCGGAAGGTCTGCTGATGCGATTTGCAACCTCAAC 311
QY 121 CTGCTGGAAGTGACTATTTTGGCTGCTGAGTTTCTGATCACAAGATCAGGTGTGG 180
Db 312 CTGCTGGAAGTGACTATTTTGGCTGCTGAGTTTCTGATCACAAGATCAGGTGTGG 371
QY 181 CTGATCTCTTAAACCCATTGTGAACAGATTAGAGGCCAAGCAGCTGTGTGTTAAG 240
Db 372 CTGATCTCTTAAACCCATTGTGAACAGATTAGAGGCCAAGCAGCTGTGTGTTAAG 431
QY 241 TTTGTGTGAAATCTTTTCGCTGACCAACACACAACTCAA 282
Db 432 TTTGTGTGAAATCTTTTCGCTGACCAACACAACTCAA 473

RESULT 4
AAA42150/C
ID AAA42150 standard; cDNA; 241 BP.
XX
XX AAA42150;
XX
XX 21-AUG-2000 (first entry)
DT
XX

DE Human secreted expressed sequence tag SEQ ID NO:890.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; nootropic; antiproliferative;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
OS
XX WO200021990-A1.
PN
XX 20-APR-2000.
PD
XX 15-OCT-1999; 99WO-US24205.
PF
XX 15-OCT-1998; 98US-0104435.
PR
XX (GEMY) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
PI WPI; 2000-317937/27.
DR
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 354; 618pp; English.
PS
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemokinetic; proliferative; immunomodulatory; haematopoietic;
CC chemotactic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antiproliferative; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
XX Sequence 241 BP; 59 A; 58 C; 58 G; 66 T; 0 other;
SQ

Query Match 48.9%; Score 138; DB 21; Length 241;
Best Local Similarity 100.0%; Pred. No. 9.1e-36;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 CTCGAGTTTCTGATCACAAGATCAGGTGTGGTGGATCTCTCTAAACCATTTGTG 204
Db 241 CTCGAGTTTCTGATCACAAGATCAGGTGTGGTGGATCTCTCTAAACCATTTGTG 182

QY 205 AAACAGATTAGAGGCAAGCACGTTGTTGAAGTTTGTGTTGAAATCTTTCGGCT 264
 |||||
 Db 181 AAACAGATTAGAGGCAAGCACGTTGTTGAAGTTTGTGTTGAAATCTTTCGGCT 122
 |||||
 QY 265 GACCACACAACTCCAA 282
 |||||
 Db 121 GACCACACAACTCCAA 104
 |||||

RESULT 5

ABK84074
 ID ABK84074 standard; cDNA; 3643 BP.

AC ABK84074;

DT 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #645.

DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

PF 03-OCT-2000; 2000US-237189P.

PR (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity

XX Claim 1; SEQ ID No 645; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating

CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3643 BP; 1038 A; 815 C; 863 G; 927 T; 0 other;

Query Match 24.3%; Score 68.4; DB 24; Length 3643;
 Best Local Similarity 56.5%; Pred. No. 4.9e-12;
 Matches 148; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 24 CAAATCCAGATCGTGGATGACACCCAGGAGGCAATTTGAAGTTCCACAAGAGCTCCTGG 83
 |||||

Db 864 CAACATCCTTCTTCGTGATAACACTGTACAAGCTTTCAAAGTCAATAAATGATCATCAGG 923
 |||||

QY 84 GAAGTGCTGCTGGATGCAGTTTGGCAACCACTCAACCTCGTGGAGGTGACTATTTTGG 143
 |||||

Db 924 GCAAGTCCTTGTGGATGTCGCTTCAAGCATCTAGATTTGACTGAGCAGGACTATTTTGG 983
 |||||

QY 144 CCTCGAGTTTCTCTGATCA---CAAAAAGATCACGGTGTGGTGGATCTCTAAAACCCAT 200
 |||||

Db 984 TTTACAGTTGGCTGATGATTCACAGATAACCCAGGTGGTGGATCCAAACAACCAAT 1043
 |||||

QY 201 TGTGAACAGATTAGAAAGCCCAAGCACGTTGTTGTTAAGTTTGGTGAATTCCTTCC 260
 |||||

Db 1044 AAGGAGCAGCTAAAGAGAGATCTCCTTACAGTTTGAACCTTTAGAGTCAAAATTTTGT 1103
 |||||

QY 261 GCCTGACACACACACTCCAA 282

Db 1104 AAGTGACCCCAACAGTTACAA 1125

RESULT 6

AAH33845

ID AAH33845 standard; cDNA; 659 BP.

XX AAH33845;

DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:901.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

DR P-PSDB; AAG74414.

DT 26-SEP-2001 (first entry)
XX Human protein tyrosine phosphatase (PTPase) DNA #2.
XX
KW Protein tyrosine phosphatase; PTPase; human; nervous system; dementia;
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; autism;
KW Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania;
KW cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo;
KW ankylosing spondylitis; Sjogren's syndrome; asthma; atopic dermatitis;
KW diabetes mellitus; inflammatory disorder; pancreatitis; sarcoidosis; ss;
KW allogeneic transplant rejection; multiple myeloma; Burkitt's lymphoma;
KW leukaemia; cancer; neoplasm.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 230..1339
FT CDS /*tag= a
FT /product= "Human PTPase #2"
XX
XX WO200153530-A1.
PN 26-JUL-2001.
XX
XX 17-JAN-2001; 2001WO-US01563.
XX
XX 18-JAN-2000; 2000US-0176306.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Shi Y, Ruben SM;
PI
XX WPI: 2001-442265/47.
DR P-PSDB; AAU0486.
XX
XX New human protein tyrosine phosphatase polypeptide useful for treatment
PT of disorders including disorder of the nervous system -
PT
XX Claim 1; Page 282-283; 290pp; English.
XX
XX The sequence represents DNA encoding a human protein tyrosine phosphatase
CC (PTPase) polypeptide. PTPase proteins and nucleic acids are useful for
CC preventing, treating or ameliorating a medical condition. The medical
CC conditions are disorders of the central and peripheral nervous system
CC (e.g. Alzheimer's disease, Huntington's disease, Parkinson's disease,
CC Tourette's syndrome, obsessive compulsive disorder, schizophrenia, mania,
CC dementia and autism), cardiovascular disorders (e.g. atherosclerosis and
CC myocarditis), autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, ankylosing spondylitis, Addison's disease, Sjogren's syndrome,
CC vitiligo, asthma, atopic dermatitis and diabetes mellitus), inflammatory
CC disorders (e.g. pancreatitis, sarcoidosis and allogeneic transplant
CC rejection) and cancers and neoplasms (e.g. leukaemia, Burkitt's lymphoma
CC and multiple myeloma). The presence or absence of a mutation in the
CC nucleic acid encoding the protein allows for diagnosis of or
CC susceptibility to a pathological condition. The sequences are also useful
CC in screening for agonists and antagonists of PTPase activity.
XX
XX Sequence 2872 BP; 840 A; 575 C; 601 G; 856 T; 0 other;
SQ

Query Match 23.1%; Score 65.2; DB 22; Length 2872;
Best Local Similarity 54.1%; Pred. No. 5.1e-11;
Matches 133; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 35 TGCTGGATGACACCCAGGAGGCAATTGAAGTTCCACAAGAGCTCCTGGGAAGGTGCTGC 94
DB 54 TGCTCGAGGACCCGAGGAGGAGCTCGCGGACCTCGCGAAGACATCCCAAGGCCAGGATTGT 113
QY 95 TGGATGCACTTTGCAACCACTCAACCTCGTGGAGGTGACTATTGCGCTCGAGTTTC 154
DB 114 TTGATCAGATTGTGTACCACTTGGACCTGTGTGGAACAGATTACTTTGCGCTCCAGTTCC 173
QY 155 CTGATCACAAAAAGATCAGCGTGTGGTGGATCTCTCTAAACCCATTGTGAAACAGATTA 214

Db 174 TCGACTCTGCCAGGTTGGCGCACTGGCTGGATCATGCCAAACCATAAAAAGCAGATGA 233
QY 215 GAAGGCCAAAGCACGTTGTTGAAGTTTGGTGGAAATCTTTCCGCGCTGACCACACAC 274
Db 234 AAATTGGACCTGCTATGCTTTACACTTTCGAGTTAATACTATCTTCAGAACCAACA 293
QY 275 AACCTCC 280
Db 294 ACCTTC 299
RESULT 9
ABN93400
ID ABN93400 standard; cDNA; 3313 BP.
XX
AC ABN93400;
XX
DT 23-JUL-2002 (first entry)
XX
DE Human gene GS915528 coding sequence from chromosome 9q31-34 #2.
XX
KW Human; gene; chromosome 9q31-34; lipoprotein metabolism disorder;
KW cholesterol transport disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200071710-A2.
PD 30-NOV-2000.
XX
XX 25-MAY-2000; 2000WO-FR01426.
PF
XX 25-MAY-1999; 99ER-0006587.
PR 16-JUN-1999; 99US-0139450.
XX
XX (AVET) AVENTIS PHARMA SA.
PA
XX Denefle P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C;
PI
XX WPI: 2001-025161/03.
DR P-PSDB; ABP37974.
XX
XX New nucleic acid derived from human chromosome 9, used e.g. for
PT diagnosis and drug screening, derived from genes implicated in
PT disorders of lipoprotein metabolism -
XX
XX Claim 1; Page 175-176; 269pp; French.
PS
XX The present sequence is the coding sequence for a human gene from
CC chromosome 9q31-34. This sequence is likely to be involved in diseases
CC of plasmatic lipoprotein metabolism, e.g. the reverse transport of
CC cholesterol.
XX
SQ Sequence 3313 BP; 922 A; 673 C; 719 G; 998 T; 1 other;
Query Match 23.1%; Score 65.2; DB 23; Length 3313;
Best Local Similarity 54.1%; Pred. No. 5.5e-11;
Matches 133; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 35 TGCTGGATGACACCCAGGAGGCAATTGAAGTTCCACAAGAGCTCCTGGGAAGGTGCTGC 94
DB 85 TGCTCGAGGACCCGAGGAGGAGCTGAGCGTGGACCTGCCGAAACATGCCAAGGCCAGGATTGT 144
QY 95 TGGATGCACTTTGCAACCACTCAACCTCGTGGAGGTGACTATTGCGCTCGAGTTTC 154
DB 145 TTGATCAGATTGTGTACCACTTGGACCTTGTGGAACAGATTACTTTGCGCTCCAGTTCC 204
QY 155 CTGATCACAAAAAGATCAGCGTGTGGCTGGATCTCTCTAAACCCATTGTGAAACAGATTA 214
DB 205 TCGACTCTGCCAGGTTGGCGACTGGCTGGATCATGCCAACCCATATAAAGCAGATGA 264
QY 215 GAAGGCCAAAGCACGTTGTGTTAAGTTTGGTGGTGAATCTTTCCGCGCTGACCACACAC 274

Qy	155	CTGATCAAAAAGATACACGGTGTGGCTGTATCTCTAAACCCTTGTGAACACAGATTA	214
Db	9096	GTGATGTGTACAGCCAGAAAGAACTGGCTGGACCCCTCCAAAGGAGATCAAGAGCAGATCC	9155
Qy	215	GAAGGCCAAAAGCAGCTGTGTTTAAAGTTTGGTGTGAAATTCCTTCCGCTGACACACA	273
Db	9156	GGAGTGAGTGGCTGTGTGTTTGGGAGGTGGGTAGCCCTTGGAAATTTTGCTTCACA	9214
RESULT 11			
ABN93399			
ID	ABN93399	standard; cDNA; 3106 BP.	
XX	XX		
AC	ABN93399;		
XX	XX		
DT	23-JUL-2002	(first entry)	
XX	XX		
DE	Human gene GS915528	coding sequence from chromosome 9q31-34 #1.	
XX	XX		
KW	Human; gene; chromosome 9q31-34; lipoprotein metabolism disorder;		
KW	cholesterol transport disorder; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	W0200071710-A2.		
XX	XX		
PD	30-NOV-2000.		
XX	XX		
PF	25-MAY-2000; 2000WO-FR01426.		
XX	XX		
PR	25-MAY-1999; 99FR-0006587.		
PR	16-JUN-1999; 99US-0139450.		
XX	XX		
PA	(AVET) AVENTIS PHARMA SA.		
XX	XX		
PI	Denefle P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C		
XX	XX		
DR	WPI; 2001-025161/03.		
DR	P-PSDB; ABP37973.		
XX	XX		
PT	New nucleic acid derived from human chromosome 9, used e.g. for		
PT	diagnosis and drug screening, derived from genes implicated in		
PT	disorders of lipoprotein metabolism -		
XX	XX		
PS	Claim 1; Page 174-175; 269pp; French.		
XX	XX		
CC	The present sequence is the coding sequence for a human gene from		
CC	chromosome 9q31-34. This sequence is likely to be involved in diseases		
CC	of plasmatic lipoprotein metabolism, e.g. the reverse transport of		
CC	cholesterol.		
XX	XX		
SQ	Sequence 3106 BP; 877 A; 603 C; 671 G; 955 T; 0 other;		
Query Match	21.5%;	Score 60.6;	DB 23;
Best Local Similarity	53.6%;	Pred. No. 1.8e-09;	Length 3106;
Matches 126;	Conservative 0;	Mismatches 109;	Indels 0;
Gaps			
Qy	46	ACCAGGAGGCATTTGAAGTTCCACAAAGAGCTCTCGGAAGGTGCTGCTGGATGCAGTT	105
Db	1	ACCGAAGTGAGCGGTGGACCTGCCGAACATGCCAAGGCCAGGATTTGTTGACAGATT	60
Qy	106	TGCACACACCTCAACCTCGTGGAGGTGACTATTTTGGCCTCGAGTTTCCTGTATCACAAA	165
Db	61	GTGTACACATTTGGACCTTGTGGAACAGAGATTACTTTGGCCTCCAGTTCTCTGACTCTGCC	120
Qy	166	AAGATCACGCTGTGGCTGGATCTCTTAAACCCATGTGTAACACATTAGAAGGCCAAG	225
Db	121	CAGTTCCGACACTGGCTGGATCATGTCGAACCCATAAAGACGATGAAATTTGGACT	180
Qy	226	CAGTTGTTTAAAGTTTGTGTGAAATTCCTTCCGCTCGACACACACACTCC	280
Db	181	GCTTATGCTTTTACACTTTTCGAGTTAAATACTATTTTCAGAACCAACAACTTC	235

```

RESULT 12
ABQ72602
ID ABQ72602 standard; cDNA: 621 BP.
XX
AC ABQ72602;
XX
DT 03-SEP-2002 (first entry)
XX
DE Human MDDT encoding cDNA SEQ ID NO 154.
XX
KW Human; MDDT; disease detection and treatment molecule polynucleotide;
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW hepatotropic; antiinflammatory; antipsoriatic; cytosstatic; anti-HIV;
KW antiatheric; antianemic; antiasthmatic; antiatherosclerotic; antigout;
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200240715-A2.
XX
PD 23-MAY-2002.
XX
PF 06-SEP-2001; 2001WO-US27628.
XX
PR 06-SEP-2000; 2000US-230505P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230865P.
PR 06-SEP-2000; 2000US-230988P.
PR 06-SEP-2000; 2000US-230989P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daiffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
WPI; 2002-527544/56.
DR P-PSDB; ABP51385.
XX
PT Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
PT e.g. AIDS
XX
PS Claim 1; Page 393; 618pp; English.
XX
CC The invention relates to an isolated human disease detection and
CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX
Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
screening a compound for effectiveness in altering expression of a target
polynucleotide comprising. Oligonucleotides and antibodies are useful for
detecting MDDT in a sample or for assessing toxicity of a test compound,
in a diagnostic test for a condition or a disease associated with the
expression of MDDT in a biological sample, for detecting (I) in a sample,
and for purifying (I) from a sample. A composition comprising (I), an
agonist or antagonist is useful for treating a disease or condition
associated with decreased or increased expression of functional MDDT.
(I) or (II) are useful for diagnosing, treating or preventing disorders
associated with aberrant expression of MDDT, where the disorders are
selected from a cell proliferative disorder such as arteriosclerosis,
cirrhosis, hepatitis, psoriasis, and cancer and an
autoimmune/inflammatory disorder such as AIDS, Addison's disease,
allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
rheumatoid arthritis. (II) are useful for creating knockin humanised
animals or transgenic animals to model human diseases, in somatic or
germline gene therapy, to generate a transcript image of a tissue or cell
type, for detecting differences in the chromosomal location due to
translocation or inversion among normal, carrier or affected individuals
and as hybridisation probes for mapping naturally occurring genomic
sequences.
XX
SQ Sequence 621 BP; 141 A; 162 C; 189 G; 129 T; 0 other;
Query Match 20.5%; Score 57.8; DB 24; Length 621;
Best Local Similarity 52.2%; Pred. No. 7e-09;
Matches 128; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
Qy 24 CAAATCCAGATGCTGGATGACACCCAGGAGGCAATTTGAAGTTCCACAAAGAGCTCCTGG 83
Db 202 CACCGTGGCGTGTGGACGACAGCGAGTACACCTGCACCATCCAGAGATGCCAAAGG 261
Qy 84 GAAGTGCTGCTGGATGACGATTTGCCAACCACTCAACCTCGTGGAGGCTGACTATTTGG 143
Db 262 CCAGTACCTGTTTGACCTCTTTGCCACCATCTGAACCTACTTGAGAAGACTATTTTGG 321
Qy 144 CCTCGAGTTTCTGATACAAAAAGATCACGGTGGTGGATCTCTCTAAACCCATTGT 203
Db 322 TATCCGCTTTGTAGACCCAGATAGCAGCGGCAATGGCTGGAATTTACAAAGCTGTGCT 381
Qy 204 GAAACAGATTAGAGGCCAAAGACGCTGTTGTTTAAGTTTGTGTGTAATTTCTTTCCGCC 263
Db 382 GAAACAATTGAGATCCAGCCTCCATTCCACCATGTGCTTCGCTGTGAAGTTTATCCTGC 441
Qy 264 TGACC 268
Db 442 AGACC 446
RESULT 13
ABL60869
ID ABL60869 standard; cDNA: 2846 BP.
XX
AC ABL60869;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human cytoskeleton-associated protein (CSAP)-3 cDNA (ID: 7091536CBL).
XX
KW Cytoskeleton-associated protein; CSAP; antiarteriosclerotic; human;
KW hepatotropic; antiinflammatory; antipsoriatic; cytosstatic; virucide;
KW anticonvulsant; cerebroprotective; neurotropic; neuroprotective; gene;
KW antiparkinsonian; transgenic; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 212..1924
/*tag= a
/product= "CSAP-3"
XX
PN WO200242330-A2.
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 15:33:49 ; Search time 8.57256 Seconds
(without alignments)
10088.332 Million cell updates/sec

Title: US-09-555-342B-1_COPY_151_432

Perfect score: 282

Sequence: 1 ccttcaggaaaactcgtgctc.....ctgaccacacacactccaa 282

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2.6/ptodata/1/ina/PCrUS_COMB.seq.*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37.2	13.2	562	US-09-404-879A-147	Sequence 147, App
2	31.2	11.1	35100	US-08-770-379-18	Sequence 18, Appl
3	31.2	11.1	35100	US-08-757-669A-18	Sequence 18, Appl
4	31.2	11.1	35100	US-09-230-371A-18	Sequence 18, Appl
5	31	11.0	2680	US-09-063-035-1	Sequence 1, Appli
6	31	11.0	56516	US-08-996-306-1	Sequence 1, Appli
7	31	11.0	58516	US-09-338-907-1	Sequence 1, Appli
8	31	11.0	58516	US-09-218-207-1	Sequence 1, Appli
9	31	11.0	58520	US-09-338-907-179	Sequence 179, App
10	31	11.0	58520	US-09-218-207-179	Sequence 179, App
11	30.8	10.9	9775	US-09-647-540A-1	Sequence 1, Appli
12	30.8	10.9	9775	US-09-647-540A-3	Sequence 3, Appli
13	30.8	10.9	9775	US-09-647-540A-5	Sequence 5, Appli
14	30.8	10.9	9775	US-09-647-540A-7	Sequence 7, Appli
15	30.2	10.7	750	US-09-446-504-2	Sequence 2, Appli
16	30.2	10.7	750	US-09-712-266-2	Sequence 2, Appli
17	30.2	10.7	989	US-09-446-504-41	Sequence 41, Appl
18	30.2	10.7	989	US-09-712-266-41	Sequence 41, Appl
19	29.8	10.6	1082	US-09-484-970B-109	Sequence 109, App
20	29.4	10.4	12752	US-08-459-146-1	Sequence 1, Appli
21	29.4	10.4	12752	US-08-459-065-1	Sequence 1, Appli
22	29.2	10.4	1011	US-09-105-537-26	Sequence 26, Appl
23	29.2	10.4	1184	US-09-242-095-3	Sequence 3, Appli
24	29.2	10.4	36778	US-09-105-537-5	Sequence 5, Appli
25	28.4	10.1	2749	US-08-344-695-1	Sequence 1, Appli
26	28.4	10.1	8648	US-09-415-946-2	Sequence 2, Appli
27	28.4	10.1	17341	US-09-415-946-1	Sequence 1, Appli

c 28	28.2	10.0	3991	4	US-08-506-296B-3	Sequence 3, Appli
c 29	28.2	10.0	11958	4	US-09-134-246-8	Sequence 8, Appli
c 30	28	9.9	1698	4	US-09-059-769-20	Sequence 20, Appl
c 31	27.8	9.9	387	4	US-09-124-671-16	Sequence 16, Appl
c 32	27.8	9.9	515	3	US-08-758-662-5	Sequence 5, Appli
c 33	27.8	9.9	1537	4	US-09-556-877-57	Sequence 57, Appl
c 34	27.8	9.9	1537	4	US-09-620-412C-57	Sequence 57, Appl
c 35	27.8	9.9	2225	4	US-09-276-531-109	Sequence 109, App
c 36	27.8	9.9	3600	4	US-09-657-042A-3	Sequence 3, Appli
c 37	27.8	9.9	5010	3	US-09-161-244-1	Sequence 1, Appli
c 38	27.6	9.8	462	4	US-09-643-597-253	Sequence 253, App
c 39	27.6	9.8	604	4	US-09-123-912-90	Sequence 90, Appl
c 40	27.6	9.8	604	4	US-09-643-597-90	Sequence 90, Appl
c 41	27.6	9.8	2007	4	US-09-643-597-153	Sequence 153, App
c 42	27.6	9.8	3026	4	US-09-149-476-317	Sequence 317, App
c 43	27.6	9.8	3054	4	US-09-149-476-194	Sequence 194, App
c 44	27.4	9.7	612	2	US-08-308-952-13	Sequence 13, Appl
c 45	27.4	9.7	612	4	US-09-124-141-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-404-879A-147
; Sequence 147, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-147

Query Match	13.2%	Score 37.2;	DB 4;	Length 562;
Best Local Similarity	53.4%;	Pred. No. 0.0022;		
Matches	78;	Conservative	0;	Mismatches 68;
Indels	0;	Gaps	0;	
QY	43	GACACCCAGGAGGCATTTGAAGTTCACAAAGAGCTCCTGGGAAGTGCTGCTGGATGCA	102	
Db	175	GATGCAGAGCTGGAGTTTGCANTCCAGCCAAATACAACTGGAAACAGCTTTTGTATCAG	234	
QY	103	GTTTGCACCACTCAACCTCGTGGAGTGACTATTTTGGCTCGAGTTTCTGTATCAC	162	
Db	235	GTGTGTAAGACATATCGGCTCCGGGAAGTGTGTACTTTTGGCTCCACTATGTGGATAAT	294	
QY	163	AAAAAGATCACGGTGTGGCTGGATCT	188	
Db	295	AAAGGATTTCCTACCTGGCTGAAGCT	320	

RESULT 2
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF

Db	28704	AAGGAGGGAGCTCTGCATGAAGAGGCTAAATGTTATTGAGACACAGCTCTCACCAA	28645
Qy	223	AAGCAGTGTGTTTAAAGTTTGTGGTGAATTCITTCGGCTTGACCACACAACTCCAA	282
Db	28644	CAGCCGGTGACGCAAGGGTCCGGGAGTGTGAATCTGTGACGGCAACACATTTCAA	28585

```

RESULT 5
US-09-063-035-1
; Sequence 1, Application US/09063035
; Patent No. 6160091
; GENERAL INFORMATION:
; APPLICANT: PEUKERT, Karen; HAENEL, Frank; and EILERS,
; APPLICANT: Martin
; TITLE OF INVENTION: Myc-binding zinc finger proteins,
; TITLE OF INVENTION: their preparation and their use
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.1
; SOFTWARE: Wordperfect version 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,035
; FILING DATE: 21-APR-1998
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2680 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..159
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..2571
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2572..2680
US-09-063-035-1
Query Match 11.0%; Score 31; DB 4; Length 2680;
Best Local Similarity 59.8%; Pred. No. 0.83;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 80 CTGGGAAGTCTGCTCGATGCAGTTTGCAACACCCTCAACCTCGTGGGAAGGTGACTATT 139
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2264 CTGTGAAGCAAGTCAGGAGAAGACCCCAACACTCACATCCTCTAGCGCTGTGACTCCT 2323
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 140 TTGCCCTCGAGTTTCCTGATCACAAAA 166
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2324 GTGGGGACAAGTTCTGTGATGCCAACA 2350
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-08-996-306-1
; Sequence 1, Application US/08996306
; Patent No. 5945522
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel

```

LOCATION: 18196..18265
FEATURE: Exon 3
NAME/KEY: 23717..23832
LOCATION: Exon 4
NAME/KEY: 25571..25660
LOCATION: SEQ ID43
NAME/KEY: 34216..34234
LOCATION: FEATURE: SEQ ID25
NAME/KEY: 34469..34515
LOCATION: FEATURE: SEQ ID52
NAME/KEY: compl(34625..34645)
LOCATION: FEATURE: Exon 5
NAME/KEY: 34669..34759
LOCATION: FEATURE: Exon 6
NAME/KEY: 40688..40846
LOCATION: FEATURE: Exon 7
NAME/KEY: 48070..48193
LOCATION: FEATURE: Exon 8
NAME/KEY: 50182..54523
LOCATION: FEATURE: SEQ ID65
NAME/KEY: 51149..51168
LOCATION: FEATURE: SEQ ID59
NAME/KEY: 51448..51494
LOCATION: FEATURE: SEQ ID68
NAME/KEY: compl(51482..51499)
LOCATION: FEATURE: SEQ ID44
NAME/KEY: 51596..51613
LOCATION: FEATURE: SEQ ID26
NAME/KEY: 51612..51658
LOCATION: FEATURE: SEQ ID53
NAME/KEY: compl(51996..52015)
LOCATION: FEATURE: polyAad signal
NAME/KEY: 54445..54450
LOCATION: US-08-996-306-1

Query Match 11.0%; Score 31; DB 2; Length 56516;
Best Local Similarity 52.8%; Pred. No. 4.9;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 35 TGCTGATGACACCCAGGAGGCGATTGAAGTTCACAAAGAGCTCCTGGGAAGGTGCTGC 94
Db 11827 TTCTGGAAACACTGTGTAAGCTTTTCAGTCCGAATAACATGATCATGTCGCAAGTCTGT 11886
QY 95 TGGATGCACTTTGCAACCACTCAACCTCGTGGGAAGGTGACTATTTTGGCCCTCGAGTTTC 154
Db 11887 TAGATGTAAGTCTGCAAGCATCTGATTTACTCTGGGCAAGACTATCTTGATTTACAGCGG 11946
QY 155 CTGATCA 161
Db 11947 CTGATCA 11953

RESULT 7
US-09-338-907-1
Sequence 1, Application US/09338907
Patent No. 6265546
GENERAL INFORMATION:

APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CPICP
CURRENT APPLICATION NUMBER: US/09/338,907
CURRENT FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 56516
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: promoter
LOCATION: 1629..1870
OTHER INFORMATION: identification method Proscan
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2000
OTHER INFORMATION: potential start codon
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2031..2033
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11694..14332
OTHER INFORMATION: Tyr Phos
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
FEATURE:
NAME/KEY: allele
LOCATION: 12057..12103
OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complemen
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
FEATURE:
NAME/KEY: allele
LOCATION: 13657..13703
OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complemen
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon 2
FEATURE:
NAME/KEY: exon
LOCATION: 23717..23832
OTHER INFORMATION: exon 3
FEATURE:

NAME/KEY: exon
LOCATION: 25571..25660
OTHER INFORMATION: exon 4
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34216..34234
OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
FEATURE:
NAME/KEY: allele
LOCATION: 34469..34515
OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34625..34645
OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
FEATURE:
NAME/KEY: exon
LOCATION: 34669..34759
OTHER INFORMATION: exon 5
FEATURE:
NAME/KEY: exon
LOCATION: 40688..40846
OTHER INFORMATION: exon 6
FEATURE:
NAME/KEY: exon
LOCATION: 48070..48193
OTHER INFORMATION: exon 7
FEATURE:
NAME/KEY: exon
LOCATION: 50182..54523
OTHER INFORMATION: exon 8
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51149..51168
OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
FEATURE:
NAME/KEY: allele
LOCATION: 51448..51494
OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51482..51499
OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51596..51613
OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
FEATURE:
NAME/KEY: allele
LOCATION: 51612..51658
OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
FEATURE:
NAME/KEY: primer_bind
LOCATION: 51996..52015
OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 54445..54450
OTHER INFORMATION: AATAAA

QY 155 CTGATCA 161
Db 11947 CTGATGA 11953
RESULT 8
US-09-218-207-1
; Sequence 1, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018cp1
; CURRENT APPLICATION NUMBER: US/09/218,207
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 56516
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1629..1870
; OTHER INFORMATION: identification method Proscan
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1998..2000
; OTHER INFORMATION: potential start codon
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2031..2033
; OTHER INFORMATION: ATG
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11694..14332
; OTHER INFORMATION: Tyr Phos
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12057..12103
; OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13547..13564
; OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
; FEATURE:
; NAME/KEY: allele
; LOCATION: 13657..13703
; OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13962..13981
; OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
; FEATURE:

OTHER INFORMATION: AATAAA	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 1991..2008	
OTHER INFORMATION: upstream amplification primer 5-63	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 2505..2525	
OTHER INFORMATION: downstream amplification primer 5-63 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 4091..4111	
OTHER INFORMATION: downstream amplification primer 99-622	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 4528..4546	
OTHER INFORMATION: upstream amplification primer 99-622 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 5475..5495	
OTHER INFORMATION: downstream amplification primer 99-621	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 5927..5947	
OTHER INFORMATION: upstream amplification primer 99-621 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 8127..8144	
OTHER INFORMATION: downstream amplification primer 99-619	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 8560..8578	
OTHER INFORMATION: upstream amplification primer 99-619 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 11622..11639	
OTHER INFORMATION: upstream amplification primer 4-76	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 12018..12037	
OTHER INFORMATION: downstream amplification primer 4-76 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 11930..11947	
OTHER INFORMATION: upstream amplification primer 4-77	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 12339..12358	
OTHER INFORMATION: downstream amplification primer 4-77 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 12915..12932	
OTHER INFORMATION: upstream amplification primer 4-71	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 13317..13334	
OTHER INFORMATION: downstream amplification primer 4-71 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 13216..13233	
OTHER INFORMATION: upstream amplification primer 4-72	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 13617..13636	
OTHER INFORMATION: downstream amplification primer 4-72 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 13547..13564	
OTHER INFORMATION: upstream amplification primer 4-73	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 13962..13981	
OTHER INFORMATION: downstream amplification primer 4-73 , complement	

OTHER INFORMATION: downstream amplification primer 99-610	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 16463..16480	
OTHER INFORMATION: upstream amplification primer 99-610 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 17304..17324	
OTHER INFORMATION: downstream amplification primer 99-609	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 17814..17832	
OTHER INFORMATION: upstream amplification primer 99-609 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18008..18027	
OTHER INFORMATION: upstream amplification primer 4-90	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18423..18442	
OTHER INFORMATION: downstream amplification primer 4-90 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18699..18716	
OTHER INFORMATION: downstream amplification primer 99-607	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 19164..19182	
OTHER INFORMATION: upstream amplification primer 99-607 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 22589..22609	
OTHER INFORMATION: downstream amplification primer 99-602	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 23111..23129	
OTHER INFORMATION: upstream amplification primer 99-602 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 25098..25118	
OTHER INFORMATION: downstream amplification primer 99-600	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 25657..25674	
OTHER INFORMATION: upstream amplification primer 99-600 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 26537..26557	
OTHER INFORMATION: downstream amplification primer 99-598	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 27022..27040	
OTHER INFORMATION: upstream amplification primer 99-598 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 32262..32281	
OTHER INFORMATION: downstream amplification primer 99-592	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 32823..32841	
OTHER INFORMATION: upstream amplification primer 99-592 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 34215..34233	
OTHER INFORMATION: upstream amplification primer 99-217	
FEATURE:	

Query Match 11.0%; Score 31; DB 4; Length 56520;
Best Local Similarity 52.8%; Pred. No. 4.9;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 15994..16011	
OTHER INFORMATION: downstream amplification primer 99-610	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 16463..16480	
OTHER INFORMATION: upstream amplification primer 99-610 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 17304..17324	
OTHER INFORMATION: downstream amplification primer 99-609	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 17814..17832	
OTHER INFORMATION: upstream amplification primer 99-609 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18008..18027	
OTHER INFORMATION: upstream amplification primer 4-90	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18423..18442	
OTHER INFORMATION: downstream amplification primer 4-90 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 18699..18716	
OTHER INFORMATION: downstream amplification primer 99-607	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 19164..19182	
OTHER INFORMATION: upstream amplification primer 99-607 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 22585..22609	
OTHER INFORMATION: downstream amplification primer 99-602	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 23111..23129	
OTHER INFORMATION: upstream amplification primer 99-602 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 25098..25118	
OTHER INFORMATION: downstream amplification primer 99-600	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 25657..25674	
OTHER INFORMATION: upstream amplification primer 99-600 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 26537..26557	
OTHER INFORMATION: downstream amplification primer 99-598	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 27022..27040	
OTHER INFORMATION: upstream amplification primer 99-598 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 32262..32281	
OTHER INFORMATION: downstream amplification primer 99-592	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 32823..32841	
OTHER INFORMATION: upstream amplification primer 99-592 , complement	
FEATURE:	
NAME/KEY: primer_bind	
LOCATION: 34215..34233	
OTHER INFORMATION: upstream amplification primer 99-217	
FEATURE:	

Query Match

Best Local Similarity 11.0%; Score 31; DB 4; Length 56520;

Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;


```
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:
NAME/KEY: primer_bind
LOCATION: 19164..19182
OTHER INFORMATION: upstream amplification primer 99-607, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 22589..22609
OTHER INFORMATION: downstream amplification primer 99-602
FEATURE:
NAME/KEY: primer_bind
LOCATION: 23111..23129
OTHER INFORMATION: upstream amplification primer 99-602, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25098..25118
OTHER INFORMATION: downstream amplification primer 99-600
FEATURE:
NAME/KEY: primer_bind
LOCATION: 25657..25674
OTHER INFORMATION: upstream amplification primer 99-600, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 26537..26557
OTHER INFORMATION: downstream amplification primer 99-598
FEATURE:
NAME/KEY: primer_bind
LOCATION: 27022..27040
OTHER INFORMATION: upstream amplification primer 99-598, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32262..32281
OTHER INFORMATION: downstream amplification primer 99-592
FEATURE:
NAME/KEY: primer_bind
LOCATION: 32823..32841
OTHER INFORMATION: upstream amplification primer 99-592, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34215..34233
OTHER INFORMATION: upstream amplification primer 99-217
FEATURE:
NAME/KEY: primer_bind
LOCATION: 34624..34644
```

Query Match 11.0%; Score 31; DB 4; Length 56520;
Best Local Similarity 52.8%; Pred. No. 4.9;

```
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 35 TGTGTGATGACACCCAGGAGGCAATTTGAAGTTCCACAAAGAGCTCCTGGGAAGGTCTGCG 94
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11827 TTTCTGGAACACTGTGTAGCTTTTCAGTGGCAATAAATGATGATGATGCGCAAGTTCTGT 11886
QY 95 TGGATCAGCTTTGCAACCACTCAACCTCGTGGGAAGTGACTATTTTGGCCTGCGAGTTTC 154
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11887 TAGATGTAGTCTGCAAGCATCTCTGATTTTACTTGGGCAAGACTATGTTGTTTACAGCGG 11946
QY 155 CTGATCA 161
|||||
Db 11947 CTGATGA 11953
```

```
RESULT 11
US-09-647-540A-1/c
; Sequence 1, Application US/09647540A
; Patent No. 6420158
; GENERAL INFORMATION:
; APPLICANT: Ishii, Yoshitaka
; APPLICANT: Konishi, Jin
; APPLICANT: Hirasawa, Kazuaki
; APPLICANT: Okada, Hideki
; APPLICANT: Suzuki, Masanori
; FILE OF INVENTION: GENES ENCODING DESULFURIZATION ENZYMES
; CURRENT APPLICATION NUMBER: US/09/647,540A
; PRIOR APPLICATION NUMBER: PCT/JP99/01756
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: JP 310545/1998
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: JP 090387/1998
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 9775
; TYPE: DNA
; ORGANISM: Paenibacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3031)...(4410)
US-09-647-540A-1
```

Query Match 10.9%; Score 30.8; DB 4; Length 9775;
Best Local Similarity 52.3%; Pred. No. 2.1;
Matches 68; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
QY 127 GAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAGGATCAGCGTGGCTGGAT 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1164 GATTTTGGCGTGTGTCGCCGCGGATCTTGGAGCAGTTTCAATTGCTGCGCCCTT 1105
QY 187 CTCCTAAACCCATTTGTGAACAGATTAGAAGGCCAACGAGCTTGTCTTAAGTTTGTG 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1104 ATAAGCAATTCAGCGCGGATCGTCCCGATGTTCCGACATATTCCTTCAGTTCGCG 1045
QY 247 GTGAATTCCT 256
|||||
Db 1044 GTTAAACTCT 1035
```

```
RESULT 12
US-09-647-540A-3/c
; Sequence 3, Application US/09647540A
; Patent No. 6420158
; GENERAL INFORMATION:
; APPLICANT: Ishii, Yoshitaka
; APPLICANT: Konishi, Jin
; APPLICANT: Hirasawa, Kazuaki
; APPLICANT: Okada, Hideki
; APPLICANT: Suzuki, Masanori
```

```

: APPLICANT: UEMORI, Takashi
: APPLICANT: SATO, Yoshimi
: APPLICANT: FUJITA, Tomoko
: APPLICANT: MIYAKE, Kazue
: APPLICANT: MUKAI, Hiroyuki
: APPLICANT: ASADA, Kiyozo
: APPLICANT: KATO, Ikunoshin
: TITLE OF INVENTION: DNA POLYMERASE

```



```
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-446-504-2

Query Match      10.7%; Score 30.2; DB 4; Length 750;
Best Local Similarity 50.3%; Pred. No. 0.75;
Matches 74; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 98 ATGCAGTTTGCAACCAACCTCGTGGAGGTGACTATTTTGGCCTCGAGTTTCCTG 157
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 518 AGGAAGTTGAGATAAGCTAACTCTTGAAGATGAGGGATTTTGGACATCGAGGTTCAAG 577
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 ATCACAAAAGATCAGGGTGTGGCTGGATCTCCTAAACCCATTGTGAACACAGATTAGAA 217
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 578 AGGAGACAAAGAGCGCATATGGAGTCAGCTATCTCTCCGACATGTTAAAGGACTTGGAA 637
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 218 GCCCAAGCAGCAGTTGTGTTAAGTTTG 244
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 638 AGGCCGATGAAGTTACAATAAAGTTTG 664
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: December 3, 2002, 18:26:48
Job time : 156.573 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 3, 2002, 18:25:06 ; Search time 9.318 Seconds
(without alignments)
11655.134 Million cell updates/sec

Title: US-09-555-342b-1_copy_151_432
Perfect score: 282
Sequence: 1 ccttcaggaaaactgtgtc.....ctgaccacacacaactccaa 282

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues
Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.6	25.4	466	10	US-09-728-445-750
2	65.2	23.1	2872	10	US-09-906-779-3
3	54.6	19.4	3984	10	US-09-848-294-1
4	45.2	16.0	4125	12	US-10-044-090-620
5	44.8	15.9	408	10	US-09-960-352-6773
6	44.2	15.7	417	12	US-10-044-090-624
7	38.8	13.8	300	10	US-09-998-598-1581
c 8	38.6	13.7	414	10	US-09-954-456-1461
9	37.2	13.2	512	10	US-09-960-253-27
10	37.2	13.2	562	10	US-09-884-441-147
11	37.2	13.2	2930	10	US-09-960-253-156
12	37.2	13.2	3044	10	US-09-880-107-3718
13	37.2	13.2	3047	10	US-09-864-864-329
14	37.2	13.2	3115	10	US-09-925-299-123
15	37	13.1	439	10	US-09-925-300-104
c 16	36.4	12.9	1372	10	US-09-765-111A-13
c 17	36.4	12.9	2334	10	US-09-765-111A-1
c 18	36.4	12.9	2523	10	US-09-765-111A-3
c 19	36.4	12.9	2596	10	US-09-765-111A-22

c 20	36.4	12.9	2625	10	US-09-765-111A-5	Sequence 5, Appli
c 21	36.4	12.9	2711	10	US-09-765-111A-24	Sequence 24, Appl
c 22	31	11.0	56516	9	US-09-853-526-1	Sequence 1, Appli
23	31	11.0	56516	10	US-09-901-484A-1	Sequence 1, Appli
24	31	11.0	56520	9	US-09-853-526-179	Sequence 179, App
25	31	11.0	56520	10	US-09-901-484A-179	Sequence 179, App
26	30.6	10.9	517	10	US-09-833-381-207	Sequence 207, App
27	30.6	10.9	517	10	US-09-833-381-208	Sequence 208, App
28	30.2	10.7	750	10	US-09-971-309-2	Sequence 2, Appli
29	30.2	10.7	989	10	US-09-971-309-41	Sequence 41, Appli
c 30	29.8	10.6	7574	10	US-09-070-927A-152	Sequence 152, App
c 31	29.6	10.5	695	10	US-09-867-701-10647	Sequence 10647, A
c 32	29.2	10.4	1011	9	US-09-860-846-26	Sequence 26, Appl
c 33	29.2	10.4	1011	10	US-09-861-289-36	Sequence 26, Appl
c 34	29.2	10.4	3957	12	US-10-005-467-1	Sequence 1, Appli
35	29.2	10.4	36778	9	US-09-860-846-5	Sequence 5, Appli
36	29.2	10.4	36778	10	US-09-861-289-5	Sequence 5, Appli
c 37	28.8	10.2	479	10	US-09-864-761-865	Sequence 865, App
c 38	28.8	10.2	4578	9	US-09-764-868-1401	Sequence 1401, Ap
39	28.8	10.2	4578	9	US-09-764-868-1402	Sequence 1402, Ap
40	28.6	10.1	392	10	US-09-833-381-1132	Sequence 1132, Ap
41	28.4	10.1	428	9	US-09-860-670-240	Sequence 240, App
c 42	28.4	10.1	790	10	US-09-880-107-416	Sequence 416, App
43	28.4	10.1	1693	10	US-09-822-849A-269	Sequence 269, App
44	28.4	10.1	4457	10	US-09-960-253-154	Sequence 154, App
45	28.4	10.1	9519	10	US-09-764-847-1056	Sequence 1056, Ap

ALIGNMENTS

RESULT 1
US-09-728-445-750
; Sequence 750, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; TITLE OF INVENTION: Animals
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 750
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(466)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-750

Query Match	25.4%	Score 71.6;	DB 10;	Length 466;
Best Local Similarity	57.3%;	Pred. No. 8.1e-15;		
Matches 150;	Conservative	0;	Mismatches 109;	Indels
				Gaps 1;
QY	24	CAAAATCCAGATGCTGGATGACACCCAGGAGGATTTGAAGTTCCACAAAGAGCTCCTGG	83	
Db	118	CAACATCTTCTCTGGATAACACTGTACAGGCTTTCAGAGTTAACAACATGATCAGGG	177	
QY	84	GAAAGTGCTGCTGGAGTCAGTTGCAACCCACCTCAACCTCGTGGAGGTGACTATTTTGG	143	
Db	178	GCAAGTCTGTTGGATATAGTCTTCAAGCATCTTGATTTGACTGACGAGACTATTTGG	237	
QY	144	CCTCGAGTTTCCCT---GATCACAAAAGATCAGGGTGTGCTGGATCTCCTAAAACCCAT	200	
Db	238	TTTACAGTTGGTGAGGATTCACAGATAACCCCAAGTGTGCTGGATCCCAACCAACCAAT	297	

Db 515 AAAAGATCGAGTGCACAAAGTGATACTCTCGATGGATCAGAAATATACCTGTGATGTAGAG 574

QY 70 CAAAGAGCTCCCTGGGAAGGTGCTGCTGGATGCAAGTTTGCACACCACTCAACCTCGTGGAA 129

Db 575 AAACGCTCAGAGCAAGCTGTTGATAAAGTGTGGAACACTTGAACCTTGCTAGAG 634

QY 130 GGTGACTATTTTGGCTCGAGTTTCTGTATACACAAAAGATCAGCGTGGGTGGATC 187

Db 635 AAAGACTACTTTGGGCTTACGTATCGATGCTGAAACCCAGAAAGTGGTTGGACC 692

RESULT 5

US-09-960-352-6773

Query Match 15.9%; Score 44.8; DB 10; Length 408;

Best Local Similarity 51.2%; Pred. No. 1.1e-05;

Matches 131; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 29 TCCAGATGCTGGATGACACCCAGGAGGCAATTTGAAGTTCCACAAGAGCTCCTGGGAAG 88

Db 111 TCCGCTTTTAGATGCTTGGTACAGACCTTTAAAGTTAAACAACAAGATACCGGTCAAG 170

QY 89 TCGTCTGTGATCGAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTGGCTCG 148

Db 171 TTCTCTGGACATGGCTTACACTACCTGGGTGTGACTGAGAAGGAGTATTTTCGGTTAC 230

QY 149 AGTTTCCCTGATCACAAAAGATCACGGTG---TGGCTGGATCTCTTAAACCACTTGTGA 205

Db 231 ACATGGGTGATGACTCACTGGACTCTCTAGATGGCTTGAACCAAGCAACCACTCAGGA 290

QY 206 AACAGATTAGAAGGCCAACAGCAGTGTGTGTTAAGTTTGTGGTGAATTTCTTTCGCCCTG 265

Db 291 AGCAGTTAAAGGAGGTTTCCCTCTACCTGATCCCTGATTTTCGAGTAAGATTTTATACCTG 350

QY 266 ACCACACAACTCCA 281

Db 351 ATCCCAACACACTGCA 366

RESULT 6

US-10-044-090-624

Query Match 15.9%; Score 44.8; DB 10; Length 408;

Best Local Similarity 51.2%; Pred. No. 1.1e-05;

Matches 131; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 29 TCCAGATGCTGGATGACACCCAGGAGGCAATTTGAAGTTCCACAAGAGCTCCTGGGAAG 88

Db 111 TCCGCTTTTAGATGCTTGGTACAGACCTTTAAAGTTAAACAACAAGATACCGGTCAAG 170

QY 89 TCGTCTGTGATCGAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTGGCTCG 148

Db 171 TTCTCTGGACATGGCTTACACTACCTGGGTGTGACTGAGAAGGAGTATTTTCGGTTAC 230

QY 149 AGTTTCCCTGATCACAAAAGATCACGGTG---TGGCTGGATCTCTTAAACCACTTGTGA 205

Db 231 ACATGGGTGATGACTCACTGGACTCTCTAGATGGCTTGAACCAAGCAACCACTCAGGA 290

QY 206 AACAGATTAGAAGGCCAACAGCAGTGTGTGTTAAGTTTGTGGTGAATTTCTTTCGCCCTG 265

Db 291 AGCAGTTAAAGGAGGTTTCCCTCTACCTGATCCCTGATTTTCGAGTAAGATTTTATACCTG 350

QY 266 ACCACACAACTCCA 281

Db 351 ATCCCAACACACTGCA 366

RESULT 7

US-09-998-598-1581

Query Match 15.7%; Score 44.2; DB 12; Length 417;

Best Local Similarity 57.7%; Pred. No. 1.8e-05;

Matches 79; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGATGATGACAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTT 141

Db 96 GGACAAAGTCTGCTGATGATAAAGTGTGTGAACACTTGAACCTTGTAGAGAAAGACTACTTT 155

QY 142 GGCCTCGAGTTTCTGATCACAATAAGATCAGCGTGGCTGGATCTCTTAAACCACTT 201

Db 156 GGGCTTACGTATCGAGATGCTGAAAACCAAGAAATTTGGTTGGACCTCTCTAAGGAATA 215

QY 202 GTGAACACAGATTAGAG 218

Db 216 AAAAAACAGGTTTGAAG 232

US-09-998-598-1581

Sequence 1581, Application US/09998598

Patent No. US20020150922A1

GENERAL INFORMATION:

APPLICANT: Stolk, John A.

APPLICANT: Xu, Jiangchun

APPLICANT: Chenault, Ruth A.

APPLICANT: Meagher, Madelein Joy

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.561

CURRENT APPLICATION NUMBER: US/09/998,598

CURRENT FILING DATE: 2001-11-16

NUMBER OF SEQ ID NOS: 2606

SOFTWARE: Corixa Invention Disclosure Database

SEQ ID NO 1581

LENGTH: 300

TYPE: DNA

ORGANISM: Homo sapiens

US-09-998-598-1581

Query Match 13.8%; Score 38.8; DB 10; Length 300;

Best Local Similarity 56.2%; Pred. No. 0.0011;

Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGATGATGACAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTT 141

Db 26 GGACAAAGTGTATTGACAAAGTGTGTGAACCACTCAATCTCTTGGAGAAAGACTACTTT 85

QY 142 GGCCTCGAGTTTCTGATCACAATAAGATCAGCGTGGCTGGATCTCTTAAACCACTT 201

Db 86 GGACTTTTGTTCAGGAAGCCCTGAGCAGAAAACTGGTTAGATCTCTCTAAGGAATA 145

QY 202 GTGAACACAGA 211

Db 146 AAGAGACAAA 155

RESULT 8

US-09-954-456-1461/c

Sequence 1461, Application US/09954456

Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C

FILE REFERENCE: 2002-01-09

CURRENT APPLICATION NUMBER: US/10/044,090

NUMBER OF SEQ ID NOS: 850

SOFTWARE: PERL Program

SEQ ID NO 624

LENGTH: 417

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

OTHER INFORMATION: Incyte ID No. US20020137081A1 347941.1

NAME/KEY: unsure

LOCATION: 355-356

OTHER INFORMATION: a, t, c, g, or other

US-10-044-090-624

Query Match 15.7%; Score 44.2; DB 12; Length 417;

Best Local Similarity 57.7%; Pred. No. 1.8e-05;

Matches 79; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGATGATGACAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTT 141

Db 96 GGACAAAGTCTGCTGATGATAAAGTGTGTGAACACTTGAACCTTGTAGAGAAAGACTACTTT 155

QY 142 GGCCTCGAGTTTCTGATCACAATAAGATCAGCGTGGCTGGATCTCTTAAACCACTT 201

Db 156 GGGCTTACGTATCGAGATGCTGAAAACCAAGAAATTTGGTTGGACCTCTCTAAGGAATA 215

QY 202 GTGAACACAGATTAGAG 218

Db 216 AAAAAACAGGTTTGAAG 232

US-10-044-090-624

Sequence 1581, Application US/09998598

Patent No. US20020150922A1

GENERAL INFORMATION:

APPLICANT: Stolk, John A.

APPLICANT: Xu, Jiangchun

APPLICANT: Chenault, Ruth A.

APPLICANT: Meagher, Madelein Joy

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.561

CURRENT APPLICATION NUMBER: US/09/998,598

CURRENT FILING DATE: 2001-11-16

NUMBER OF SEQ ID NOS: 2606

SOFTWARE: Corixa Invention Disclosure Database

SEQ ID NO 1581

LENGTH: 300

TYPE: DNA

ORGANISM: Homo sapiens

US-09-998-598-1581

Query Match 13.8%; Score 38.8; DB 10; Length 300;

Best Local Similarity 56.2%; Pred. No. 0.0011;

Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 82 GGAAGAGTCTGCTGATGATGACAGTTTGCACACCACTCAACCTCGTGGAAAGTGACTATTTT 141

Db 26 GGACAAAGTGTATTGACAAAGTGTGTGAACCACTCAATCTCTTGGAGAAAGACTACTTT 85

QY 142 GGCCTCGAGTTTCTGATCACAATAAGATCAGCGTGGCTGGATCTCTTAAACCACTT 201

Db 86 GGACTTTTGTTCAGGAAGCCCTGAGCAGAAAACTGGTTAGATCTCTCTAAGGAATA 145

QY 202 GTGAACACAGA 211

Db 146 AAGAGACAAA 155

RESULT 8

US-09-954-456-1461/c

Sequence 1461, Application US/09954456

Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C

FILE REFERENCE: 2002-01-09

CURRENT APPLICATION NUMBER: US/10/044,090

NUMBER OF SEQ ID NOS: 850

SOFTWARE: PERL Program

SEQ ID NO 624

LENGTH: 417

TYPE: DNA

```

; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1461
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1461

```

Query Match 13.7%; Score 38.6; DB 10; Length 414;
Best Local Similarity 56.8%; Pred. No. 0.0015;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy	35	TGCTGGATGACCCAGAGGCCATTTGAAGTTCACAAGAAGCTCTGGAAAGTGCTCGC	94
Db	211	TTCTGGATGTACTGATGTTTAGTGTGGACITGC AAAAAGCCAAGGCAAGAGTTCG	152
Qy	95	TGATGTCAGTTTGC AACCACCTCAACCTCGTGG AAGGTGACTATTTTTGGCC TCGAGTTTC	154
Db	151	TTGATCAGATTATGATCATCACCTGGATCTTATCAAAGTGGCTATTTTGATCTGAGATTTTA	92
Qy	155	CTGAT	159
Db	91	TGGAT	87

```

RESULT 9
US-09-960-253-27
; Sequence 27, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-27

```

Query Match 13.2%; Score 37.2; DB 10; Length 512;
Best Local Similarity 53.4%; Pred. No. 0.005;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

3

Qy	43	GACACCACGAGGCA	TTTGAAGTTCCACA	AAGAGCTCTCGGGAAGT	GCTGCTGGATGCA	102
Db	112	GATCGAGCTGGAGT	TTTGCAATCCAGCCA	ATAAATGGAACACGCT	TTTTGATCAG	171
Qy	103	GTTTGGAAACCACT	CAACCTCGTGGAGG	TGACTATTTTGGCCCT	CGAGTTTCTCGATCAC	162
Db	172	GTGTAAGAACTAT	CGCCCTCCGGAAGT	GTGTACTTTGGCCCT	CATATGTGGATAAT	231
Qy	163	AAAAGATCAGGTG	TGGCTGGATCT			188
Db	232	AAAGATTTCCTAC	TCTCGCTGGAAGCT			257

RESULT 10

```

US-09-884-441-147
; Sequence 147, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: fastSEQ for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-884-441-147

```

Query Match	13.2%	Score 37.2;	DB 10;	Length 562;
Best Local Similarity	53.4%;	Pred. No. 0.0053;		
Matches 78;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0;

[illegible]

RESULT 11

```

US-09-960-253-156
; Sequence 156, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF LUNG CANCER
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-156

```

Query Match 13.2%; Score 37.2; DB 10; Length 2930;
Best Local Similarity 53.4%; Pred. No. 0.013;

Search completed: December 3, 2002, 22:25:28
Job time : 16.318 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 3, 2002, 16:31:10 ; Search time 303.618 seconds
(without alignments)
15042.359 Million cell updates/sec

Title: US-09-555-342b-1_COPY_151_432
Perfect score: 282
Sequence: 1 ccttcagagaaactgtgtc.....ctgaccacacacaactccaa 282

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	282	100.0	859	9	AL537488
2	282	100.0	995	13	BM549907
3	278.8	98.9	450	9	AI690075
4	276	97.9	441	9	AI937921
5	274.4	97.3	433	9	AI298784
6	221.8	78.7	398	9	AI879532

c	7	215.4	76.4	482	12	BF011995
c	8	213.8	75.8	478	14	W82071
	9	212.2	75.2	629	12	BF021124
	10	208.2	73.8	472	12	BF940343
	11	207.4	73.5	1185	12	BF3227626
c	12	193.6	68.7	354	12	BE835601
	13	178	63.1	631	12	BG019331
	14	177.2	62.8	1002	12	BG323892
	15	171	60.6	314	14	T05026
	16	170.4	60.4	612	14	BQ383992
	17	170.4	60.4	619	14	BQ397802
	18	156.4	55.5	298	12	BE816579
	19	150	53.2	570	9	AI091437
	20	132	46.8	570	12	BF195016
c	21	120	42.6	179	14	BQ305056
c	22	104.6	37.1	121	9	AI298786
	23	102.8	36.5	327	14	T07137
	24	97.8	34.7	679	12	BF772813
	25	97.8	34.7	752	13	BI763945
	26	97.8	34.7	788	12	BG718448
	27	97.8	34.7	971	14	BM910128
	28	97	34.4	741	9	AL555132
	29	92.2	32.7	517	9	AI528026
	30	92.2	32.7	626	13	BI697755
	31	92.2	32.7	635	10	BB654641
	32	92.2	32.7	652	10	BB663505
	33	92.2	32.7	898	14	BQ944256
	34	90.6	32.1	518	13	BI319861
	35	83.6	29.6	685	9	AU168701
	36	82.8	29.4	488	10	BB840554
	37	81.8	29.0	325	12	BE816570
	38	72.8	25.8	416	12	BF914765
	39	72.4	25.7	350	13	BJ062492
	40	71.6	25.4	465	9	AA415407
	41	71.6	25.4	552	9	AI595558
	42	70.4	25.0	639	13	BM290978
	43	69	24.5	368	9	AA758137
	44	69	24.5	555	14	BM849444
	45	69	24.5	584	14	BM782015

ALIGNMENTS

RESULT 1
AL537488 859 bp mRNA linear EST 13-FEB-2001
LOCUS AL537488 LTI_FLO13_FBrnl Homo sapiens cDNA clone CS0DF025YJ01 5
DEFINITION prime, mRNA sequence.
ACCESSION AL537488
VERSION AL537488.1 GI:12800981
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 859)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF025YJ01"
/clone_lib="LTI_FLO13_FBrnl"
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
/lab_host="DH10B"

/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 224 a 237 c 220 g 169 t 9 others
ORIGIN

Query Match 100.0%; Score 282; DB 9; Length 859;
Best Local Similarity 100.0%; Pred. No. 3.9e-81;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
|||||
Db 318 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 377
|||||
QY 61 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGGATGCAAGTTCGCAACCACTCAAC 120
|||||
Db 378 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGGATGCAAGTTCGCAACCACTCAAC 437
|||||
QY 121 CTCGTGGAAGTGACTATTTTGGCCTCGAGTTTCTGATCACAAAAAGATCACGGTGTGG 180
|||||
Db 438 CTCGTGGAAGTGACTATTTTGGCCTCGAGTTTCTGATCACAAAAAGATCACGGTGTGG 497
|||||
QY 181 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACCTTTGTTTAAAG 240
|||||
Db 498 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACCTTTGTTTAAAG 557
|||||
QY 241 TTGTGCTGAATCTTTCGGCTGACACACACACTCAA 282
|||||
Db 558 TTGTGCTGAATCTTTCGGCTGACACACACACTCAA 599
|||||

RESULT 2
BM549907
LOCUS BM549907 995 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6558078 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5443230
5', mRNA sequence.

ACCESSION BM549907
VERSION BM549907.1 GI:18785692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 995)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: L1CM1920 row: g column: 07
High quality sequence stop: 736.
Location/Qualifiers
1. .995

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5443230"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRT; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 245 a 304 c 267 g 178 t 1 others
ORIGIN

Query Match 100.0%; Score 282; DB 13; Length 995;
Best Local Similarity 100.0%; Pred. No. 4.3e-81;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
|||||
Db 368 CCTTCAGGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 427
|||||
QY 61 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGGATGCAAGTTCGCAACCACTCAAC 120
|||||
Db 428 GAAGTTCCACAAGAGCTCCTGGAGGTGCTCGGATGCAAGTTCGCAACCACTCAAC 487
|||||
QY 121 CTCGTGGAAGTGACTATTTTGGCCTCGAGTTTCTGATCACAAAAAGATCACGGTGTGG 180
|||||
Db 488 CTCGTGGAAGTGACTATTTTGGCCTCGAGTTTCTGATCACAAAAAGATCACGGTGTGG 547
|||||
QY 181 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACCTTTGTTTAAAG 240
|||||
Db 548 CTGGATCTCTAAACCCATTTGGAACAGATTAGAGGCCAAGACACCTTTGTTTAAAG 607
|||||
QY 241 TTGTGCTGAATCTTTCGGCTGACACACACACTCAA 282
|||||
Db 608 TTGTGCTGAATCTTTCGGCTGACACACACACTCAA 649
|||||

RESULT 3
AI690075/c
LOCUS AI690075 450 bp mRNA linear EST 16-DEC-1999
DEFINITION tx27d07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2270797 3'
similar to WP:H05G16.1 CELL1568 BAND 4.1 FAMILY ;, mRNA sequence.

ACCESSION AI690075
VERSION AI690075.1 GI:4901369
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 673 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1. .450

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2270797"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"

/lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "
 BASE COUNT 106 a 111 c 109 g 124 t
 ORIGIN

Query Match 98.9%; Score 278.8; DB 9; Length 450;
 Best Local Similarity 99.3%; Pred. No. 3e-80;
 Matches 280; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCTTCAGGAAACTCTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
 Db 439 CCTCAAGGAAACTCTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 380
 QY 61 GAAGTTCACAAAGAGCTCCTGGGAAGTGTCTGCTGATGCAGTTTGCACACCACTCAAC 120
 Db 379 GAAGTTCACAAAGAGCTCCTGGGAAGTGTCTGCTGATGCAGTTTGCACACCACTCAAC 320
 QY 121 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTG 180
 Db 319 CTCGTGGAAGGTGACTATTTTGGCCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTG 260
 QY 181 CTGGATCTCTAAACCCATTGTTGAGCCATGTAACAGATAGAAAGCCAAAGACAGTTGTTGTAAG 240
 Db 259 CTGGATCTCTAAACCCATTGTTGAGCCATGTAACAGATAGAAAGCCAAAGACAGTTGTTGTAAG 200
 QY 241 TTGTGCTGAAATCTTTCGGCTGACACACACCACTCAA 282
 Db 199 TTGTGCTGAAATCTTTCGGCTGACACACCACTCAA 158

RESULT 4
 AI937921/c
 LOCUS
 DEFINITION
 au54f07.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
 IMAGE:2518597 3' similar to WP:H05G16.1 CELL1568 BAND 4.1 FAMILY ;,
 mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)

JOURNAL
 COMMENT
 Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40UP from Gibco
 High quality sequence stop: 429.
 Location/Qualifiers

FEATURES
 source
 1..441
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone="IMAGE:2518597"
 /clone_lib="Schneider fetal brain 00004"
 /sex="male"
 /tissue_type="frontal lobe"
 /dev_stage="5 months post-conception"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);
 Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
 prepared from human fetal brain tissue. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence:
 5'-GAGAGAGAGAGCTCAAGATCCTTAATTAATATATCCCCCCCCC-3'
 and 3' adaptor sequence:
 5'-GAGAGAGAGACTCGAGTTTTTTTTTTT-3'. The library was
 size-selected for >0.5 kb inserts and has an average
 insert size estimated at 1.2 kb. This library was
 constructed using the CAP-trapper method for full-length
 enrichment and has not undergone amplification. Library
 was constructed by Dr. Claudio Schneider (LNCIB-Area
 Science Park, Trieste, Italy)."
 BASE COUNT 105 a 110 c 100 g 126 t
 ORIGIN

Query Match 97.9%; Score 276; DB 9; Length 441;
 Best Local Similarity 100.0%; Pred. No. 2.5e-79;
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GGAAGAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTT 66
 Db 441 GGAAGAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTT 382
 QY 67 CCACAAAGAGCTCCTGGGAAGTGTCTGCTGGATGCAAGTTCGAACCACTCAACCTCGTG 126
 Db 381 CCACAAAGAGCTCCTGGGAAGTGTCTGCTGGATGCAAGTTCGAACCACTCAACCTCGTG 322
 QY 127 GAAGTGTACTATTTTGGCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGGCTGGAT 186
 Db 321 GAAGTGTACTATTTTGGCCTCGAGTTTCCTGATCACAAAAGATCACGGTGTGGCTGGAT 262
 QY 187 CTCCTAAACCCATTGTGAACAGATTAGAGCCCAAGACAGCTTTGTTTAAGTTTGTG 246
 Db 261 CTCCTAAACCCATTGTGAACAGATTAGAGCCCAAGACAGCTTTGTTTAAGTTTGTG 202
 QY 247 GTGAAATCTTTCCGCTGACACACACCACTCAA 282
 Db 201 GTGAAATCTTTCCGCTGACACACACCACTCAA 166

RESULT 5
 AI298784/c
 LOCUS
 DEFINITION
 qm91b09.x1 NCI CGAP Lu5 Homo sapiens cDNA clone IMAGE:1896089 3'
 similar to WP:H05G16.1 CELL1568 ;, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 433)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 702 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 359.

FEATURES

source

```

1. .433
  Location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:1896089"
    /clone_lib="NCI_CGAP_Lu5"
    /tissue_type="carcinoid"
    /lab_host="DH10B"
    /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
    modified polylinker; 1st strand cDNA was prepared from a
    neuroendocrine lung carcinoid, and was then primed with a
    Not I - oligo(dT) primer. Double-stranded cDNA was ligated
    to Eco RI adaptors (Pharmacia), digested with Not I and
    cloned into the Not I and Eco RI sites of the modified
    pT7T3 vector. Library is normalized. Library was
    constructed by Bento Soares and M. Fatima Bonaldo. "
```

```

BASE COUNT      105 a      98 g      99 c      98 g      119 t
ORIGIN
Query Match      97.3%; Score 274.4; DB 9; Length 433;
Best Local Similarity 99.6%; Pred. No. 8.2e-79;
Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 7 GGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATTTGAAAGTT 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 433 GGAAACTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGAGGCATTTGAAAGTT 374

QY 67 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGAGTTGCAACCACTCAACCTCGTG 126
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 373 CCACAAAGAGCTCCTGGGAGGTGCTGCTGGATGAGTTGCAACCACTCAACCTCGTG 314

QY 127 GAAGTGTACTATTTGGCCCTCGAGTTTCTCGATCACAAAAGATCACGGTGGCTGGAT 186
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 313 GAAGTGTACTATTTGGCCCTCGAGTTTCTCGATCACAAAAGATCACGGTGGCTGGAT 254

QY 187 CTCCTAAACCCATGTTGAAACAGATTAGAAGCCAAACAGCTGTTGTTTAAAGTTTGTG 246
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 253 CTCCTAAACCCATGTTGAAACAGATTAGAAGCCAAACAGCTGTTGTTTAAAGTTTGTG 194

QY 247 GTGAATTTCTTCCGCTGCTGACCAACACACAACTCCAA 282
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 193 GTGAATTTCTTCCGCTGCTGACCAACACACAACTCCAA 158
```

RESULT 6

A1879532/c

```

LOCUS      A1879532      398 bp      mRNA      linear      EST 23-AUG-1999
DEFINITION au54c11.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone
            IMAGE:2518580 3' similar to WP:H05G16.1 CELL568 BAND 4.1 FAMILY ;,
            mRNA sequence.
```

ACCESSION A1879532

VERSION A1879532.1 GI:55533590

KEYWORDS EST.

SOURCE human.

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

REFERENCE

AUTHORS

```

1 (bases 1 to 398)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,X., Wylie,T., Waterston,R. and Wilson,R.
```

TITLE

JOURNAL

COMMENT

```

Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
```

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand

Seq primer: -40UP from Gibco

High quality sequence stop: 245.

FEATURES

source

```

1. .398
  Location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2518580"
    /clone_lib="Schneider fetal brain 00004"
    /sex="male"
    /tissue_type="frontal lobe"
    /dev_stage="5 months post-conception"
    /lab_host="DH10B"
    /note="Organ: brain; Vector: pBluescript SK (Stratagene);
    Site.1: SstI; Site.2: XhoI; Double-stranded cDNA was
    prepared from human fetal brain tissue. 5' and 3'
    adaptors were used in cloning as follows: 5' adaptor
    sequence:
    5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCCCCCCCCCC-3'
    and 3' adaptor sequence:
    5'-GAGAGAGAGACTCGAGTGTGTTTTTTTTTTT-3'. The library was
    size-selected for >0.5 kb inserts and has an average
    insert size estimated at 1.2 kb. This library was
    constructed using the CAP-trapper method for full-length
    enrichment and has not undergone amplification. Library
    was constructed by Dr. Claudio Schneider (LNCIB-Area
    Science Park, Trieste, Italy)."
```

```

BASE COUNT      99 a      98 c      86 g      115 t
ORIGIN
Query Match      78.7%; Score 221.8; DB 9; Length 398;
Best Local Similarity 99.1%; Pred. No. 1.3e-61;
Matches 223; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 58 TTTGAAGTTCACAAAGAGCTCTGGGAGGTGCTGCTGGATGCGAGTTTGCACACACCTC 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 398 TTTGAAGTTCACAAAGAGCTCTGGGAGGTGCTGCTGGATGCGAGTTTGCACACACCTC 339

QY 118 AACCTCGTGAAGGTGACTATTTTGGCCCTCGAGTTTCTCGATCACAAAAGATCACGGTG 177
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 338 AACCTCGTGAAGGTGACTATTTTGGCCCTCGAGTTTCTCGATCACAAAAGATCACGGTG 279

QY 178 TGGCTGGATCTCCTAAACCCATTTGAAACAGATTAGAAGCCAAACAGCTGTTGTTGT 237
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 278 TGGCTGGATCTCCTAAACCCATTTGAAACAGATTAGAAGCCAAACAGCTGTTGTTGT 219

QY 238 AAGTTGTGTGTCAAATCTTCCGCTGACCAACACACAACTCCAA 282
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 218 AAGTTGTGTGTCAAATCTTCCGCTGACCAACACACAACTCCAA 174
```

RESULT 7

BF011995/c

LOCUS BF011995

DEFINITION ux52g02.y1 Soares_NKWD_mandible Mus musculus cDNA clone

IMAGE:3513938 5' similar to TR:Q9Y4F1 Q9Y4F1 CDEP. ;, mRNA

sequence.

ACCESSION BF011995

VERSION BF011995.1 GI:10712270

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 482)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

ACCESSION BF021124
 VERSION BF021124.1 GI:10752456
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 629)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other_ESTs: ux52g02.y1
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1396818
 Possible reversed clone: similarity on wrong strand
 High quality sequence stop: 429.
 FEATURES
 Location/Qualifiers
 1..629
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3513938"
 /clone_lib="Soares_NKWM_mandible"
 /tissue_type="mandible"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGGCGCGCTTAATTTTGTGTGT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo. "
 BASE COUNT 140 a 205 c 171 g 113 t
 ORIGIN
 Query Match 75.2%; Score 212.2; DB 12; Length 629;
 Best Local Similarity 84.7%; Pred. No. 2.4e-58;
 Matches 238; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 1 CCTTCAGGAAACTCTGTCTCAATAAATCCAGATGCTGGATGACACACCCAGGAGCATTT 60
 Db 258 CCTTCAGGAAACTTATGACTGCAAAATTCAGATGCTGGATGACACACCCAGGAGCATTT 317
 QY 61 GAAGTTCCACAAAGAGCTCTGGGAAGGTGCTGCTGGATGACATTTGCAACACCACTCAAC 120
 Db 318 GAAGTTCCACAAAGAGCGCCAGGAGGTCTGTTGACGCCGTGTGCAACCACTCAAC 377
 QY 121 CTGCTGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTATCACAACCAAGATCAGCGGTGG 180
 Db 378 CTGCTGAAGGTGACTACTTCGGCCCTGGAGTTCCCTGACACAGGAAGATCGTGTGTGG 437
 QY 181 CTGGATCTCTTAAACCCATTGTGAACACAGATTAGAGGCCAAGCAGCTTGTGTTAAG 240
 Db 438 CTGGATCTCTTGAAGCCATTGTGAAGCAGATTTCGAAGCCGAGCAGCGTGTGTTAA 497
 QY 241 TTTGTGGTGAATTTCTTCCGCTCGACCACTCAAC 281
 Db 498 TTTGTGGTGAATTTCTTCCCTCCGACCACTCAAC 538
 RESULT 10
 BF940343
 LOCUS 472 bp mRNA linear EST 30-MAR-2001
 DEFINITION nad95e02.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:343343 3',
 similar to TR:Q94F1 Q94F1 CDEP.;, mRNA sequence.
 ACCESSION BF940343
 VERSION BF940343.1 GI:12357663
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 472)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
 Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 378.
 FEATURES
 Location/Qualifiers
 1..472
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3433443"
 /clone_lib="NCI-CGAP_Ov18"
 /tissue_type="fibrotheoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGGCGCGCTTAATTTTGTGTGT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "
 BASE COUNT 105 a 145 c 142 g 80 t
 ORIGIN
 Query Match 73.8%; Score 208.2; DB 12; Length 472;
 Best Local Similarity 98.6%; Pred. No. 4.2e-57;
 Matches 210; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CCTTCAGGAAACTCTGTCTCAATAAATCCAGATGCTGGATGACACACCCAGGAGCATTT 60
 Db 260 CCTTCAGGAAACTCTGTCTCAATAAATCCAGATGCTGGATGACACACCCAGGAGCATTT 319
 QY 61 GAAGTTCCACAAAGAGCTCTGGGAAGGTGCTGCTGGATGACATTTGCAACCACTCAAC 120
 Db 320 GAAGTTCCACAAAGAGCTCTGGGAAGGTGCTGCTGGATGACATTTGCAACCACTCAAC 379
 QY 121 CTGCTGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTATCACAACCAAGATCAGCGGTGG 180
 Db 380 CTGCTGAAGGTGACTATTTTGGCCCTCGAGTTTCTGTATCACAACCAAGATCAGCGGTGG 439
 QY 181 CTGGATCTCTTAAACCCATTGTGAACAGATT 213
 Db 440 GTGGATCTTCTCAACCCATTGTGAACAGATT 472
 RESULT 11
 BG327626
 LOCUS 1185 bp mRNA linear EST 27-FEB-2001
 DEFINITION 602426G20F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564636 5',
 mRNA sequence.
 ACCESSION BG327626
 VERSION BG327626.1 GI:13134064
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 1185)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI280 row: 9 column: 05
High quality sequence stop: 564.

FEATURES
Location/Qualifiers
1. .1185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4564636"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 378 a 377 c 271 g 159 t

Query Match 73.5%; Score 207.4; DB 12; Length 1185;
Best Local Similarity 94.2%; Pred. No. 1.3e-56;
Matches 259; Conservative 0; Mismatches 11; Indels 5; Gaps 4;
QY 13 CTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCCACAA 72
Db 2 CTCGTGTCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTTGAAGTTCCACAA 61
QY 73 AGAGCTCCTGGGAAGTGTCTGGATGACAGTTTGAACACAC-CTCAACCTCGTGGGAAG 131
Db 62 AGAGCTCCTGGGAAGTGTCTGGATGACAGTTTGAACACACCTCAACCTCGTGGACGG 121
QY 132 TGACTATTTT-GGCCCTCGAGTTTCCCTGATCACAAAAGATCACGGTGTGGCTGATCTCC 190
Db 122 TGACTATTTTGGCCCTCGAGTTTCCCTGATCACAAAAGATCACGGTGTGGCTGATCTCC 181
QY 191 TAAACCCCATTTGTGAA-ACAGATTAGAAGCCCAAGACAGTTGTTT--GTTAAGTTTGTGG 247
Db 182 TAGAACCCCATTTGTGAACACAGATTAGAAGCCCAAGACAGTTTGTTCGTTAAGTTTGTGG 241
QY 248 TGAATTCCTTCGCCCTGACACACACAACTCCAA 282
Db 242 TGAATTCCTTCGCCCTGACACACAACTCCCA 276

RESULT 12
LOCUS BE835601/c 354 bp mRNA linear EST 22-SEP-2000
DEFINITION RC6-FN0027-110700-011-F03 FN0027 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE835601
VERSION BE835601.1 GI:10267979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 354)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

TITLE Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
COMMENT Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC6-FN0027-110
700-011-F03&t3=2000-07-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 43
High quality sequence stop: 353.

FEATURES
Location/Qualifiers
1. .354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FN0027"
/dev_stage="Adult"
/note="Organ: prostate_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 98 a 89 c 81 g 86 t

Query Match 68.7%; Score 193.6; DB 12; Length 354;
Best Local Similarity 98.0%; Pred. No. 2.2e-52;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 83 GGAAGTGCTGCTGGATGTCAGTTTGCAACCCACCTCAACCTCGTGGAGGTGACTATTG 142
Db 354 GGATGCTGCTGGATGTCAGTTTGCAACCCACCTCAACCTCGTGGAGGTGACTATTG 295
QY 143 GCCTCGAGTTTCTGATCACAAGATCAGCGTGTGGTGGATCTCCTAAAACCCCATG 202
Db 294 GCCTCGAGTTTCTGATCACAAGATCAGCGTGTGGTGGATCTCCTAAAACCCCATG 235
QY 203 TGAACAGATTAGAAGCCCAAGACAGTTGTTGTTAAGTTTGGTGAATCTTCTCCG 262
Db 234 TGAACAGATTAGAAGCCCAAGACAGTTGTTGTTAAGTTTGGTGAATCTTCTCCG 175
QY 263 CTGACCACACACAACTCCAA 282
Db 174 CTGACCACACAACTCCAA 155

RESULT 13
LOCUS BG019331
DEFINITION BG019331 631 bp mRNA linear EST 24-JAN-2001
IMAGE:4083497 5' similar to TR:Q914F1 Q914F1 CDEP. ;, mRNA
sequence.
ACCESSION BG019331
VERSION BG019331.1 GI:12475410
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 631)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,T., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Willson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by C. Hashimoto, Ph.D. in the laboratory of K.
Cho, Ph.D. DNA Sequencing by: Washington University Genome
Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 431.

FEATURES

```

1. 1001
  /organism="Xenopus laevis"
  /db_xref="taxon:8355"
  /clone="IMAGE:4083457"
  /clone_lib="Cho Li treated
  /tissue_type="gastrula, Li-
  /lab_host="DH10B"
  /note="vector: p Bluescript
  EcoRI; 1st strand was primer
  primer, double-stranded cDNA
  Eco RI sites of pBluescript
  by C. Hashimoto, Ph.D. in
  (Department of Developmental
  of California, Irvine)."
190 a 135 c 153 g 153 t

```

BASE COUNT	190 a	135 c	153 g	153 t

Query Match 63.1%; Score 178; DB 12; Length 631;
Best Local Similarity 77.0%; Pred. No. 3.9e-47;
Matches 217; Conservative 0; Mismatches 65; Indels

Qy	1	CTTTCAGGAAATCTGTTGTCCTCAATAATCCAGATGCTGGATGACACCCAGGAGGCATTT	60
Db	149	CCAAAGGGAATCTATCCATCAAGTCCAGATGCTGGATGACACCTCGAACCATT	208
Qy	61	GAAGTTCCACAAAGAGCTCCTCGGAAGGTGCTGCTGGATGCAGTTTCAACCACTCAAC	120
Db	209	GAAGTTCCCAATAGAGCCCCAGGGAAGCTCTTTGGATGCCGTTTCAGCCACTCAAT	268
Qy	121	CTCGTGAAGAGGTGACTATTTTGGCTCGAGTTTGCCTGATCACAAAAGATCACGGTGTGG	180
Db	269	CTTGTGTAAGGGGACTATTTTGGTCTTGAGTTCCAGGACCATAGAAAGATTATGGTCTGG	328
Qy	181	CTGGATCTCTCTAAACCCATTTGTGAACAGATTAGAAGGCCAAGCAGCTTGTTGTTAAG	240
Db	329	CTGGATCTTTTAAAGCCAATATCAACACAGATTAGAAGGCCCTAGAACGCAGTCTTAAG	388
Qy	241	TTTGTGGTGAAATCTTTTCCGCTGACCCACACAACTCAA	282
Db	389	TTTGTAGTCAAGTTTTTTTCCACTGTATCACACCTCAGCTGCAA	430

RESULT 14

Accession	BG323892
Definition	LOCUS
Accession	BG323892
Definition	602422921f1 NTH_MGC_14 Homo sapiens cDNA clone IMAGE:4560868 5'
Accession	BG323892
Definition	mRNA sequence.
Accession	BG323892
Definition	1002 bp linear EST 27-FEB-2001
Accession	BG323892
Definition	GI:13130329

KEYWORDS	EST.
SOURCE	human
ORGANISM	Homo s
	Eukary
	Mamma
REFERENCE	1 (b
AUTHORS	NIH-M
TITLE	Nation
JOURNAL	Unpub
COMMENT	Contac

Contact: Robert Sfrausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling H
cDNA Library Arrayed by: The I.M.
DNA Sequencing by: Incyte Genomics
Clone distribution: MGC clone di
found through the I.M.A.G.E. Cons
http://image.llnl.gov
Plate: LLCIM270 row: j column:
High quality sequence start: 10
High quality sequence stop: 612

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

location/guadalupe
1. .1002
/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="IMAGE:4560868"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site:
EcoRI; cDNA made by oligo-dT priming. D
cloned into EcoRI/XhoI sites using the f
adapter: GGCACGAG(G). Size-selected >50
insert size 1.8kb. Library constructed
the laboratory of Gerald M. Rubin (Unive
California, Berkeley) using ZAP-cDNA sy
(Stratagene) and Superscript II RT (Life
259 c 250 q 191 t

BASE COUNT	302 a	259 c	250 q	191 t

```
Query Match          62.8%; Score 177.2; DB 12; Length 1002;
Best Local Similarity 92.4%; Pred. NO. 9.3e-47;
Matches 242; Conservative 0; Mismatches 13; Indels 7; Gaps 5;
```

Qy	21	CATCAAATCCAGATGCTGGATGACACCAGGAGGCATTTGAAGTTCACAAAGAGCTCC	80
Db	9	CATCAAATCCAGATGCTGGATGACACCAGGAGGCATTTGAAGTTCACAAAGAGCTCC	65
Qy	81	TGGAAGGTGCTGCTGGATGCAGTTTGCACACACCTCAACCTCGTGGAGGTGACTATTT	140
Db	66	TGGAAGGTGCTGCTGGATGCAGTTTGCACACACCTCAACCTCGTGGAGGTGACTA-TT	123
Qy	141	TGGCTCGAGTTTCTTGATCACAAAAGATCAGGTGTGGCTGGATCTCCTAAAACCCAT	200
Db	124	TGGCTCGAGTATCCTGATCACAAAAGATCAGGTGTGGCTGGATCTCCTAAAACCCA-	182
Qy	201	TGTGAACACATTAGAGGCCAAGCACCGTTGTTGTT-AGTTTGTGGTGAATCTCTTC	259
Db	183	TGTGAACACATTAGAGGCCAAGCACCGTTGTTGTTCAAGTTTGTGGTGAATCTCTTC	242
Qy	260	CGCTGACCACACAACTCCA	281
Db	243	CGCTGACCACGCACAATCAA	264

RESULT 15
T05026

105020	T05026	314 bp	mRNA	linear	EST 30-JUN-1993
LOCUS	EST02914	Fetal brain, Stratagene	(cat#936206)	Homo sapiens	CDNA
DEFINITION	clone HFBCP32, mRNA sequence.				
ACCESSION	T05026				
VERSION	T05026.1	GI:316181			
KEYWORDS	EST.				

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 314)
TITLE Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
3,400 expressed sequence tags identify diversity of transcripts
from human brain
JOURNAL Nat. Genet. 4, 256-267 (1993)
MEDLINE 93364420
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES
source
1..314
/organism="Homo sapiens"
/db_xref="ATCC (inhost):81769"
/db_xref="taxon:9606"
/clone="HF8CP32"
/clone_lib="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
oligo-dT + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average inser size."
BASE COUNT 82 a 78 c 87 g 63 t 4 others
ORIGIN

Query Match 60.6%; Score 171; DB 14; Length 314;
Best Local Similarity 98.4%; Pred. No. 5.1e-45;
Matches 182; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 60
Db 130 CCTTCAGGAAACTCGTGCCATCAAAATCCAGATGCTGGATGACACCCAGGAGCATTT 189

Qy 61 GAAGTTCCACAAGAGCTCCTGGGAAGTGCTGCTGGATGCAAGTTTGCAACCACTCAAC 120
Db 190 GAAGTTCCACAAGAGCTCCTGGGAAGTGCTGCTGGATGCAAGTTTGCAACCACTCAAC 249

Qy 121 CTCGT-GGAAGTGACTATTTGGCTCGAGTTTCCTGATCACAAAAAGATCACGGTGTG 179
Db 250 CTCGTGGGAAGGTGACTATTTGGNCTCGAGTTTCTGATCACAAAAAGATCACGGTGTG 309

Qy 180 GCTGG 184
Db 310 GCTGG 314

Search completed: December 3, 2002, 21:24:16
Job time : 318.618 secs

THIS PAGE BLANK (USPTO)